IMPROVED USABILITY OF ELECTRONIC GOVERNMENT SERVICES FOR THE AGEING POPULATION

Dissertation

Zur Erlangung des akademischen Grades Dr. rer. nat.

Im Fach Psychologie

eingereicht an der

Mathematisch-Naturwissenschaftliche Fakultät II

Humboldt-Universität zu Berlin

von

Dipl. Wirt.-Inf. Tamas Molnar

Präsident der Humboldt-Universität zu Berlin

Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät II

Prof. Dr. Elmar Kulke

Gutachter:

1. Prof. Dr. Hartmut Wandke
2. Prof. Dr. Peter Schirmbacher
3. Prof. Dr. Iván Futó

Tag der Verteidigung: 17.02.2014

Zusammenfassung

Die Fragestellung dieser Arbeit ist ob derzeit angebotene e-Government Systeme von älteren Nutzern angenommen werden und wie solche aufgebaut werden müssen, damit diese Nutzergruppe solche Systeme als eine nützliche Alternative zu Behördengängen annimmt.

In unserer Forschung haben wir erforscht wie solche Anwendungen, welche von der Verwaltung für die gesamte Bevölkerung angeboten werden, aufgebaut werden sollten, damit diese von der gesamten Bevölkerung erfolgreich genutzt werde können.

Zur Beantwortung dieser Fragestellung wurde eine dreistufige Forschung durchgeführt, welche an das ISO 9241-210 Entwicklungsmodell angelehnt ist. Hierdurch wurden die wichtigsten Kernmerkmale von e-Government Systemen entsprechend dieser speziellen Anforderungen untersucht. Die Forschung wurde parallel in Deutschland und Ungarn in Kooperation mit dem Fraunhofer FOKUS, dem Bundesministerium des Innern, der Bundesdruckerei und der Corvinus Universität Budapest durchgeführt.

In der ersten Phase wurden die Erwartungen und Vorkenntnisse der Zielgruppe erforscht um die Eckpunkte und Prämissen festlegen zu können. Dieser Schritt hat ein umfangreiches Bild darüber ermöglicht, welche Services von Personen im Alter zwischen 60 und 75 angenommen werden, und durch welche theoretischen Verbesserungen sich die Personen angesprochen fühlen. Dies wurde anhand einer Interviewstudie mit 70 Teilnehmern anhand des Reifegradmodells der Europäischen Kommission in beiden Ländern durchgeführt. Ein weiteres Ziel war die generellen Vorkenntnisse der Zielgruppe mit interaktiven Internetanwendungen zu erfassen um in den folgenden Schritten adäquate Systeme für die Nutzertests auswählen zu können.

Diese Erkenntnisse ermöglichten in der zweiten Phase die fundierte Auswahl einer Anwendung, welche als Basis für Nutzertests genutzt werden konnte. Die Testanwendung war das AusweisApp des elektronischen Personalausweises. Diese wurde mit 75 Teilnehmern, sowie mit einer Kontrollgruppe bestehend aus 20 Studierenden der Humboldt-Universität durchgeführt. Bei diesen Tests wurden die Nutzerfehler erfasst und die Akzeptanz durch die bei IBM entwickelten ASQ Methode gemessen. Dies ermöglicht die Vergleichbarkeit der erfassten Daten mit der späteren Prototyp Phase.

Anhand der gewonnen Erkenntnisse konnte die Guideline IGUAN entwickelt werden, welche eine standardisierte Herangehensweise zur Akzeptanzsteigerung darstellt. Dieses Konzept beinhaltet neben den speziellen, an ältere Nutzer angepassten Anforderungen, einem Kriterienkatalog, sowie die Abbildung der Prozesse wodurch eine Erhöhung der Akzeptanz für Ältere ermöglicht wird.

In der dritten Phase der Forschung konnte die Guideline durch eine iterative Prototypentwicklung evaluiert und geprüft werden. Dieser Schritt kann als Proof-of-Concept angesehen werden, und ermöglicht eine allgemeine Anwendung des IGUAN bei der Entwicklung von e-Government Systemen.

Wir konnten beweisen, dass Verbesserungen beim Interface e-Government Anwendungen an die alternde Gesellschaft näher bringen, die Motivation erhöhen und das Nutzerempfinden nachhaltig verbessern. Hierfür ist jedoch ein strukturierter Entwicklungsprozess für das Interface notwendig, welches durch das von uns entwickelte Guideline ermöglicht wird.

Abstract

Our research focuses on the question of acceptance of current e-government systems by elderly users. It describes how such systems should be designed and offered for this user group in order to provide an acceptable alternative to offline processes.

We have concentrated on the factors which influence the acceptance of e-government systems. This was accomplished with the premise that these applications have to be successfully usable by every cohort of the population.

In order to answer our research question, the research was structured into three phases along the development model of the ISO 9241-210. This enabled to identify not only the main factors of acceptance, but also the expectations of elderly users. Our research was conducted in parallel in Germany and Hungary in cooperation with the Fraunhofer FOKUS, the Federal Ministry of Interior, the Bundesdruckerei and the Corvinus University Budapest.

The first phase of our research provided results about the expectations and previous experience of the users with e-government systems. This set the premises for the next phases of our research and provided us with information about the theoretical acceptance of systems by the age group. In addition, the interviews with the 70 participants designed along the maturity model of the European Commission delivered first results about the possibilities of e-government for elderly users. Our goal in this phase was to acquire information about the general experience of the age group with interactive applications so that a suitable test environment could be selected in the second phase of the research.

The results made it possible to select an application in the second phase, which was used as a model in the remaining phases. The selected application was the AusweisApp of the electronic ID card. This was tested with 75 participants and a control group consisting of 20 students of the Humboldt-University. The test measured the encountered errors and the acceptance of the system with the ASQ method developed at IBM. This enabled the comparability of the results with the later design phase.

The obtain results allowed us to develop a generalised solution, the IGUAN guideline. This guideline makes a standardised approach to the usability improvement process possible. It contains the special requirements of elderly users, and a catalogue of criteria, which helps to develop an application in line with the set requirements.

The third phase of our research was used a proof of concept for the IGUAN. The guideline was evaluated and tested with an iterative prototyping. The successful completion of this phase indicates that the IGUAN can be used to measurably increase the acceptance of e-government systems by elderly users.

We could therefore demonstrate that improvements in the interface make e-government application possible which are perceived useful and easy to use by elderly users. These improvements will measurably increase the user motivation and experience. This can however only be achieved with a structured design process, and requires a framework which takes the requirements of the elderly users into account.

.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

1. B2C – Business to Consumer interaction
2. G2C – Governemt to Consumer interaction
3. ASQ – After Scenario Questionnaire
4. ASQ_{pe} – ASQ score for the perceived easiness of the application
5. ASQ_{pt} – ASQ score for the perceived time gain of the application
6. ASQ_{pu} – ASQ score for the perceived utility of the application
7. CLS – Computer Literacy Scale
8. UML – Universal Modelling Language
9. GUI – Graphical User Interface
10. HTML- Hypertext Markup Language
11. XML – Extended Markup Language
12. PIN – Personal Input Number
13. N - Number
14. M - Mean
15. SD – Standard Deviation
16. UI – User Interface
17. EULA – End User Licence Agreement
18. ATM – Automatic Teller Machine
19. ACT-R – Adaptive Control of Thought - Rational

1. Introduction - Problems of Electronic Government Services for elderly users

It was the private sector, which first discovered the powerful capabilities of the Internet in interacting with its customers. A wide range of services has been made possible in the late 1990s, like e-banking, e-commerce or e-learning. Some of these were extremely successful, and have become accepted by large portions of the general population, becoming de facto standards for business to consumer (B2C) interaction. Others are only used by certain demographic groups, and there were and still are countless failures.

The main problem of electronic systems, be it e-government, e-commerce or other e-services, is the question of acceptance. The transition from traditional services allows to or in some cases forces the customer to take a new role in the delivery of the service, but this comes at the cost of a deeper user interaction. This interactive role the user takes, means that such systems have to fit with the expectations of the user to be considered satisfactory. This solution for this however, becomes challenging as systems become more and more complex.

1.1 Research Outline

Electronic government (or e-government) as an idea was created on the example of e-business, or electronic commerce. It is essentially the evolution of the transformation of the public sector, which, as a process has been on-going for several decades, (Lenk K., 1994) but as stated by some authors (Lenk K., 1994), has only reached its critical mass in the last ten years. An e-government system, compared to e-commerce can be defined as the “overall-transactional journey, constructed of smaller encounters between employees and customers, customers and technology, and technology and employees” (Forlizzi, 2010).

The use of IT in public administration and other branches of government (including parliaments and the judiciary) as support systems has reached a high level in industrialised countries, but there was almost no political interest in this on-going and almost invisible process of modernisation of the government for a long time. Only academics and some farsighted consultants insisted on the significance of IT in public governance and its modernisation (Snellen & van de Donk, 1998).

For a long time the idea of e-government was not generally accepted. (Lenk K. , 1998) New Public Management, which seemed to be the most important explicit movement in governmental modernisation, hardly recognised the true potential of IT for changing the work practices and the business processes in the public sector (Lenk & Traunmüller, 2002). IT was regarded as an auxiliary tool which was needed, not for its potential, but because it was the trend of time. The real breakthrough came with the idea of the "Information Society" in the mid-1990s in the United States, which was not only heralded for its potential to renew society, but which was also seen as a concept to streamline and improve the performance of the public sector by creating a true service based approach. This theory has led to countless reports, which drove governments in industrialised countries towards the idea of the information society. American and Asian examples have greatly increased the interest in the potential of e-government as one of the most important building blocks of an information society (Lenk & Traunmüller, 2002).

Although the idea itself provides great opportunities for improving government processes and interaction on the G2C level, the basic concept of electronic government remains vague. The main problem, which can be still seen, is that decision makers try to base the underlying idea of e-government on concepts based on experience with e-commerce. This may misdirect the attention of governments when trying to gear services for innovation. E-government systems are somewhat different. The underlying technology of e-commerce and e-government systems is analogous, but premises are very different. A difference with a major impact on the system design is the question of the targeted customers. E-commerce systems have a customer base with clear boundaries. E-government systems have to be offered for all members of the population. Fragmentation of the users is not an option. This issue is one of the key factors which has to be considered, when working with e-government system. This is therefore one of the main elements, which determined the direction of our research.

This statement also defines the basic methodology which can be used when dealing with usability issues of e-government services. A heuristic evaluation of such systems uncovers only shortcomings which effects the usability. An expert cannot perfectly simulate the difficulties and issues experienced by specific users, who unfamiliar with the system. De Jong and Van der Geest discussed this issue by evaluating municipal e-government web services in the Netherlands (de Jong & Lentz, 2006). Formative evaluation studies, particularly those that focus on an in-depth analysis of user problems in actual-use situations uncover more usability issues, but can lead to user group specific designs. (De Jong & Van der Geest, 2000). Kantner and Rosenbaum (Kantner & Rosenbaum, 1997) propos therefore a sequential order between the two evaluation types. Research should, in their view, first conduct an expert focused evaluation followed by user-focused testing. We also use this combination of both methods as it provides the most profound results for software usability assessment when dealing with systems, where a broad user group has to be modelled.

E-government services are particularly suited for a service-designed approach, but a key factor and challenge of many systems remain; systems are perceived less useful and less easy to use than intended (Kotamraju & van der Geest, 2012). The difficulty of governments in creating e-government systems perceived as useful by the citizens is twofold: On the one hand, the needs of service builders are prioritised before the needs of the users, which results systems, which are streamlined for governmental efficiency, but not for user-experience. On the other hand, systems are not designed around the issue of acceptance by the wide array of users. This would also include users with very low or minimal previous experience with e-services. As seen in the Capgemini report (Capgemini, Rand Europe, IDC, Sogeti and DTi, 2009) these are not isolated issues, but can be observed in almost any country in the European Union, with probably the exception of Nordic countries, where e-services are more deeply integrated into the everyday life. We therefore focus in our research on multiple EU countries, in order to obtain a universally applicable solution to the problems of e-services for elderly users.

1.2 Advantages of Electronic Government Systems

E-government enables in theory the users a previously unparalleled communication interface with the government. This not only provides both sides with an efficient tool to enhance the exchange of information, but also enables a faster processing of information for both sides. This theoretical system is therefore considered by all involved parties to be much more effective and efficient method of communication. Without e-government, external an internal interfaces of the government are complex beyond reason and impair efficacy. This leads to slow communication and cumbersome processes handling for citizens and business when governmental services are involved.

The centralised nature of governmental agencies cause centralised and location bound service points, which is considered troublesome by citizens and business in the information age. E-government offers the opportunity to offer the equal level of service for the population or businesses, time or location independent. Electronic services are not time or location dependent by definition, and can thereby offer superior service quality for parts of the population which could otherwise not be given a good access to these services with traditional methods. This core ability can be used to counteract social and demographic changes. Elderly users are one of the demographic groups, who can most profit from this location independent equal service quality.

The European Commission therefore strongly supports the distribution of e-government system in the member states. The process to affect member states to implement new systems was the on the main agenda of several initiatives in the past decade, containing strict and ambitious goals for services to be implemented. The main umbrella strategy for this was the i2010 agenda, created as a follow up for the eEurope 2005 program. It set the main goals for development of information and communication technology until the year 2010 with the help of a unified framework built as guidance for the member states. This framework included sections defining the services, which should have been electronically implemented until 2010. It did not contain rules or standards for technologies to be used in the implementation. Now it is clear that the goals were only partially achieved; several key systems were successfully implemented throughout the European Union, but some major features are still not implemented. The main challenge, which can be observed, however, is the lack of thought about the users in the last decade. The i2010 focused on the technological aspects of the systems, which could or could not be reached as set out. The new Europe 2020 agenda seems to correct this error and it currently seems that the flagship initiative, "*The Digital Agenda of Europe*", will focus on this question, driving the changes in electronic government systems with the user in mind.

This new approach and the new views about user centred e-government are as of slowly being implemented, with some national projects serving as prototypes for this new concept.

1.3 Comparison of Usage of e-Government in the Member States of the EU

The initiatives, which have been shown in the previous section, are the basis for the more detailed aims, which are scheduled to be reached in this decade. These are much more applicable compared to the very broad targets set out on the strategic level. The most important aspect of these targets is the question of e-service penetration, which is one of the key indicators of acceptance of systems. Results in later sections will show that although the penetration of the Internet is quite high in European countries, the penetration of e-government is extremely low. The methodology used to assess previous user experience and computer literacy will directly illustrate this problem in Germany and Hungary. The selection of the same population group from two member states of the European Union enabled a cross-cultural study for the impact of e-government usability for the ageing population. The choice of Hungary as a comparable test environment to Germany was made on multiple premises: The e-government distribution, the Internet penetration and the demographic structure. The centralised approach to e-government in contrast to the federal system of Germany enabled an evaluation of multiple governmental procedures in a similar environment. The selection of Hungary was also made possible by the knowledge of the Hungarian language. A further aspect, which enabled a broader view on the e-government usage was the scepticism of the users with e-services, which was observed by Abramson and Morin (Abramson & Morin, 2003) in Hungary. An example for this is the low usage of the electronic tax declaration by the citizens, especially by the elderly (Inforum, 2008).

The results of Eurostat e-government usage panel (Eurostat, 2011) also support the selection of the two countries. Although it does not create separate data records for different age cohorts in E-government usage, the data in figure 1 visualises the similarities in acceptance.

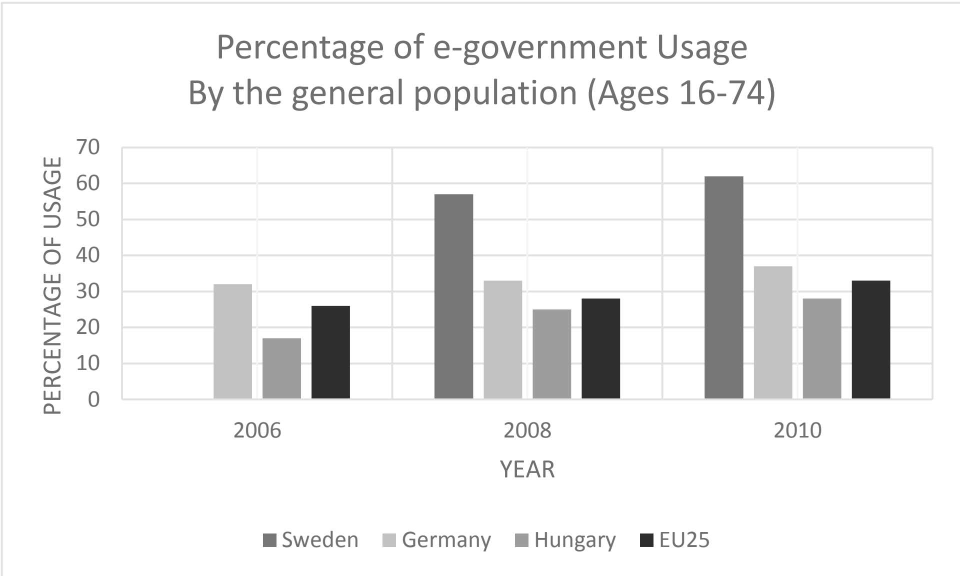


Fig. 1. - E-government Usage by the general population (Source: Eurostat 2011)

The sample size of 4000 per member state seems to be appropriate for obtaining representative results. According to Eurostat, the samples were gathered through face-to-face and telephone interviews from a heterogeneous background to represent the general population of the member states. The participation of the member states in 2006 was not mandatory, and therefore there is no data available for Sweden. In addition the differentiation between system maturity levels was not mandatory and the results therefore do not distinguish between service types. It can be seen in figure 1 that the user percentage of e-government systems has only increased marginally in the last five years throughout Europe. E-government usage in Germany and Hungary has grown only by 5% in two years, and can be considered average in the European Union. Scandinavian countries like Sweden, which are traditionally leaders in electronic services, definitely have a head start in user acceptance. The reasons for this are twofold: The low population density increases the perceived need and advantages of electronic offerings and the systems themselves can be considered better. According to Grönlund (Grönlund, 2002), the high penetration was reached, particularly in Sweden, by designing systems in accordance with the user needs and by iterative design. This is also a strong indicator, that with some thought given on usability, high levels of acceptance are possible. Without the use of user driven assessment, this diffusion level is not reachable. This is needed however for the ubiquitous services mandated by the European Commission.

A further question when dealing with e-government services is the complexity of the integrated electronic components. The maturity model for e-government systems created by the European Commission can illustrate this increasing complexity and development of e-government solutions. (Fig. 2) This model, based on the maturity levels, will be used in this thesis for the different systems, enabling the easy and proven assessment of electronic services, thereby creating a framework for the research and an option to compare systems with different components and goals.

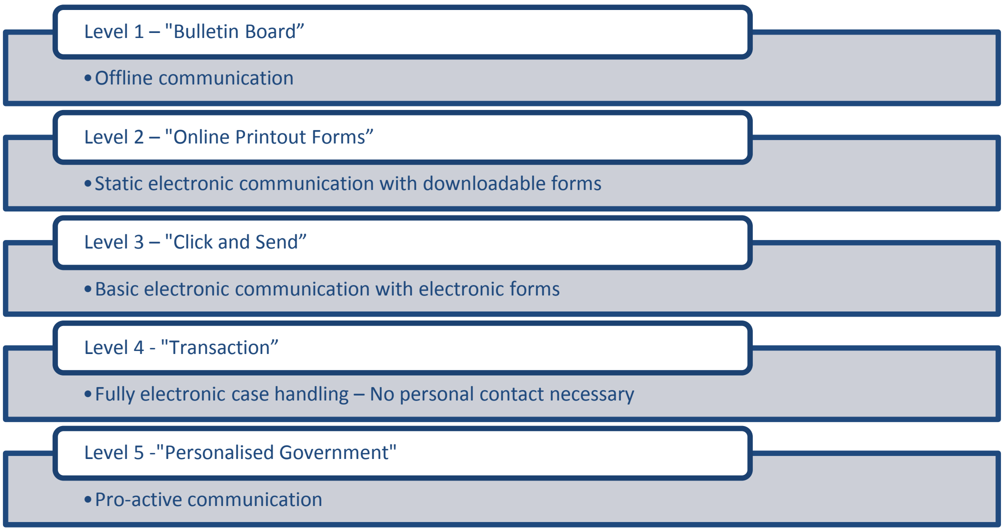


Fig. 2. - Maturity levels of e-government solutions (Source: European Commission, 2011)

A level 1 system function as an electronic "bulletin board", with only very basic information available electronically to the user. This information is offered on web page and only provides addresses or opening times. A level 2 application offers some additional benefits for the users, and offers information about needed forms or prerequisites at other agencies involved in the governmental process. This is still a static webpage, although with downloadable forms. A level 3 service is a one-way electronic solution. The users can electronically fill out forms and print them out for manual delivery. This level is advised by the European Union for several services in the i2010 agenda. A level 4 service, is the ultimate electronic solution, including fully electronic case handling, where the user can communicate with the government without personal contact. Forms can be filled out online and sent with click. Internal communication should also be fully electronic at this level, creating a completely paperless administration. This two-way electronic communication has however several pitfalls, which are discussed later.

The most recent maturity model of the European Commission include a fifth level of e-service maturity. These level 5 systems are comparable in function to the previous level, they are fully electronic without the need for paper based or face-to-face communication. In addition they are also built around the concept of proactivity. Systems designed for this level would offer the applicant services automatically based on previous communication. In other words, this level offers pro-active service delivery by offering certain social and economic rights to the citizens linked to definite previous triggers set by the applicant. As these services are currently not commonly offered and privacy issue are not fully investigated for such trigger based pro-active G2C communication, these are exempted from the research. In addition, the electronic interface for such services is not fundamentally different from level 4 systems.

1.4 Acceptance of Electronic Government Systems

We have shown earlier that e-government and e-commerce system are comparable in technical aspects, but very different in user perception. The development of e-government systems is currently approximately at a state of e-commerce systems of the early- to mid-2000s. The first euphoria is already over. Systems, like the Austrian “Bürgerkarte”, a citizen authentication card in Austria, have been introduced as the next generation authentication documents, but have failed to have a significant impact. The technology of such system was high advanced and successfully implemented, but the acceptance of the users never reached the intended levels. The Bürgerkarte was not only designed as a replacement for traditional authentication documents, but also as a platform for diffusion of the electronic signature. It failed however to gather enough interest from citizens to reach a critical mass even after years. It has been first issued in 2004 and the target date for the complete electronic authentication was set for 01-01-2008. According to the Centre for Secure Information Technology, Austria, (Zentrum für Sichere Informationstechnologie Austria, 2008) by 2006 only about 1% of the authentication was done electronically through the new system. The goals for the digital signature as a service could not be reached even two years later. Acceptance of the system lagged behind schedule and the numbers indicate that the citizens do not feel the need to use the new service. Even in 2009 the number of issued cards was about 150 000 after multiple service initiatives. This number seems to be constant since then (Österreichische Computer Geselschaft, 2012). This debacle illustrates the problems of introducing a technologically highly advanced system, without the prospective users in mind. New systems not only have to implement new services but also new opportunities to make the life easier for citizens. Against this background, the question arises: *"(Why would the citizens consider the new service superior and how can citizens be convinced…)"* transl. from (Österreichische Computer Geselschaft, 2012).

An e-government system, as any new system is considered complex and has to be accepted to be used by the target group when first encounter. A feasibility to increase the acceptance of the new system to a satisfactory level is to offer a perceived benefit for the citizens. This has to be accomplished without creating new disadvantages. E-commerce providers had to face this problem in the early- to mid-2000s. It had to be proven to customers, that their systems are better than traditional businesses. E-commerce has by definition however an enormous advantage in this regard: they can prove their superiority over traditional businesses by setting prices so that they undercut the traditional commerce, giving users a tangible benefit.

This is not possible for providers of e-government systems. Additionally, e-government systems are typically more complex than traditional e-commerce applications. Governmental processes are characteristically more difficult to model into a e-service and result a more complex structure.

The other aspect, which can be found in e-commerce, although not on this level, is the issue of privacy and the distrust in security of the systems. This problem is getting stronger with the increased complexity of the application. Low usability in such crucial systems can cause even more serious faults, as a careless click can send personal data to unauthorized parties, creating a security flaw, which can be exploited to scare users and nullify any previous acceptance of the system. Therefore, one of the main aspects of e-government applications should be a fool-proof design, which leaves little to no room for user errors. The W3B (Fittkau & Maaß Consulting, 2009) user panel for e-commerce systems in 2009 illustrated this problem. 14.4% of online shopping transactions are aborted because of the perceived high complexity and low usability of the system. If the system becomes more complex, this problem also increases.

A major difference to e-commerce is however, that governmental processes for citizens have to be always offered offline. The electronic components have to be considered an option to the traditional communication, but cannot supersede it. Therefore, it is not possible to force citizens to use e-government by eliminating the offline processing. It has to be maintained for users, who are not willing to use the new system. Laws do not only support this, but governmental services have to be equally accessible for every citizen.

An additional aspect, which is not observed in e-commerce, is also caused by the universal nature of the services provided. E-government systems have to be designed with two very distinct characteristics in mind:

1. The main drive behind e-government is the, already shown aspect of efficiency, which means that the new systems have to be seen as open and efficient. Users will only be inclined to learn and use new systems if they are either time-saving or have additional feature which are not possible offline.
2. The systems are strongly governed by external factors, not necessarily of technical nature, on both feature and system level. E-government is affected not only be privacy and relating issues, but also by laws and governmental policies.

The user perception of e-government systems is therefore much more complex than that of e-commerce systems. These system have to be offered for all citizens, and the ultimate goal is, that even the late majority and laggards (Rogers, 1962) consider these systems to be superior to offline processing.

1.5 The Ageing Population in Relation to Electronic Government Systems

As seen earlier, the understanding of usability issues of e-government has not been the highest priority of governments. Research has however shown, that elderly users would profit immensely form such systems, when communicating with the government. Results about this cohort as users of e-government systems are very limited. There have been some projects, which made superior progress in the field of general usability of e-government. On the other hand, several teams have explored problems of usability in regard of the elderly for specific applications, like mobile phones (Bruder C. , 2008), web sites (Hart T. A., 2004) (Hart, Chaparro, & Halcomb, 2008) or other electronic appliances. Very few projects conducted research on e-government specifically for the elderly, and as seen in the research of Sayago et al. (Sayago, Sloan, & Blat, 2011), most concentrate on electronic forms.

The lack of research in this direction indicates that elderly users are not considered as prime users of e-government, in contrast to the advantages for them seen in the previous chapter. Decision-makers do not understand the implications, before understanding the demographic projections and their impact governmental procedures. Elderly users, a population group defined by the WHO (World Health Organisation, 2012) as above the age 65, will become a dominant part of the population in the next decades.

A common misconception, which leads to the dismissal of this population group as users, is the claim, that older people are unwilling to use technology, such as computers. Czaja and Lee (Czaja & Lee, 2007) have disproved this stereotype and their results indicate that older people are receptive to technology. Research shows that Internet usage has strongly increased in this population group in the last five years. Current data show that 57% of individuals between 60 and 69 can be regarded as regular Internet users, and even in the age group of 70+, almost every fourth individual uses the Internet regularly (TNS Infrastest, 2011).

The higher life expectancy is increasing (Martin & Kliegel, 2005) the percentage of elderly in the whole population. In Germany in 2008, the number of people over 65 was larger than the age group under 20 (Statistisches Bundesamt, 2011). In 2025, this numbers will increase even further and the percentage of elderly will become even more dominant. It is well documented (Kuhlmey, 2006) that current elderly cannot be compared to the same age group 40 years ago. Advances in preventative medicine and research have not only caused a much better statistical health and mobility in this age group, but they have also changed their perception and views about their place in the society (Oswald & Wahl, 2010).

A second factor, which makes the current and future elderly differ from the same age group 20 years ago, is that the percentage of high education and high income in this group is strongly increasing as people, who were educated in or after the 1960s[1] become pensioners. The views of this generation are drastically different from those born before them. These elderly want to keep up with trends; they want to use the Internet (ARD/ZDF-Online-Studie, 2007). Some parts of this age group have however quite a hard time using interactive systems for the first time (Ziefle & Bay, 2003). One solution seen in some market segments is to offer special devices with fewer features for the elderly. In case of e-government systems however, this might not be a feasible solution. Research done by several groups (Bruder C. , 2008), (Hart, Chaparro, & Halcomb, 2008) with usability of other electronic devices, like mobile phones, websites, entertainment electronics or navigation systems has shown that the majority of elderly require functionality, but also require an easy to use device. It has also been shown that a majority does not accept the peculiar appearance of special senior devices (Bruder C. , 2008). This can be illustrated by the sales figures of senior phones, which have been brought to the market by some providers throughout Europe, and sold at a very bad rate, with elderly people preferring fully featured mobile phones (Bruder C. , 2008).

It seems that the need for better usability is highly important for these users, but the features (and the image) seem to have a similar value. The image of lesser abilities is considered extremely negative and elderly do not want to show their shortcomings by using devices or systems designed to look and feel simple by removing any complexity not only in the user interface, but also in function. This factor seems however to be inversely

[1] In western Europe the social revolution in the 1960s has caused that the percentage of highly educated increased in the population, in eastern Europe, the percentage of highly educated has strongly increased from the 1960s through easier access to higher education

related to the age of the user. Younger elderly despise these devices, but the practical usability features as key size and other issues seem to outweigh the bad image for users over the age of 70 (Hart T. A., 2004).

This means that understanding their unique requirements is becoming paramount in the design of e-government systems. Early studies in the field of general usability requirements for the elderly by Coyne and Nielsen (Coyne & Nielsen, 2002) already indicated that older adults (over 65) experienced about half the level of usability of younger users. Results by Mead et al. (Mead, Spaulding, Sit, & Walker, 1997) also indicate that older users have unique difficulties when using the Web. It has however been shown by Czaja and Sharit (Czaja & Sharit, 1998) or Harte et al (Hart, Chaparro, & Halcomb, 2008), that system can be built around the concept of acceptance by elderly users. This also demonstrates that e-services, and in particular e-government can be designed in accordance with this goal, and acceptance can be increased through prioritizing usability. This difference in Internet penetration also warrants the question about the difference between the elderly. It has been shown by multiple authors (Nielsen J. , 1993), (Czaja & Lee, 2007) that elderly users have to be divided into two separate groups: the young elderly and old elderly, to be able to assess their different requirements and experiences. The definition of young and old elderly will be defined later in this chapter. Behaviour and connection to technology is fundamentally different between these cohorts. This has to be considered when conducting research in this field. Psychology does not concur where the line between the two groups should be drawn.

It is important to recognize, that the older adults as group are very heterogeneous and individually very different. Ageing is a highly individualised process and with increasing age there is an increase in inter-individual differences in rate, onset and direction of change in most functions and processes (Birren & Schaie, 1996). This means that older adults vary considerably in their abilities, skills and experiences. Most researchers currently use the age of 75 as a divider, but even this line is not rigid (Martin & Kliegel, 2005). This divide can be considered dependant on the technology which is observed, and its period of mass-diffusion among the general population. As mobile phones were already common in the late 1990s, the divide can be put today as high as the age 75. Interactive Internet applications, even e-commerce services, only reached widespread diffusion by the early 2000s and may warrant therefore a divide at the age of 70. This divide can be translated to e-government, which can be considered comparable to e-commerce in many aspects. Other features shown below, like the retirement age or the diffusion of the particular technology in the work environment, also contribute to this decision.

The **old elderly** - The cohorts older than 70 are considered to be in this group. These generations seem to be less interested in new concepts and technology than the younger cohorts are. This general observation can be based on several social aspects. Members of these cohorts were born before 1940 and were already settled in the economic boom of the early 1960s, which had therefore less impact on their behavioural development. The social changes of the late 1960s also influenced most members of this group to a much lesser degree than younger generations. The percentage of higher degrees in this generation is also not as high as for those born after 1940. An aspect, which has a crucial influence on the theoretical acceptance of new interactive applications, is also deterministic – Most members of this generation reached retirement age in the early 1990s or even earlier before the political changes in Central Europe[2]. This means that even the highly educated members of these cohorts had little contact with computers in their work life, and their first exposure to this technology had to be voluntary. This aspect causes the rift between the two groups and supports the hypothesis that the divide should be set at the age of 70.

The **young elderly** - This group is composed of the cohorts between the ages of 65 and 70. Members of this group were born in the 1940s and were young in the 1950s and 1960s at a time when advances in technology were getting faster, thereby reaching the phase of diffusion at a steady pace. These generations also witnessed and profited from the rapid economic growth of the 1960s, and took active part in the social upheavals of the late 1960s, becoming mostly open-minded and curious about new ideas. The economic boom also enabled more members of this generation to acquire higher schooling than any

[2] The retirement age in Hungary was under 60 around 1990. Meaning that cohorts born before 1940 retired with 55-60 in the 1990s, without coming in “real” contact with the Internet at work.

generation before them. Lastly, these cohorts were still part of the active workforce in the late 1990s when computers became widespread in offices. Most members of this generation with at least secondary education had to work with computers later in their work life, reaching retirement age by the late 1990s or early 2000s. Their higher affinity to the Internet is also supported by the results of the TNS Infratest panel (TNS Infrastest, 2011).

This means that most members of this generation are motivated to try new, innovative concepts such as e-government, wanting to learn how to use new technologies. It is important for the younger elderly that the application is fully featured and easily usable at the same time. These attributes make it understandable, why this generation despises the use of special senior devices, which not only look peculiar, but also lack several features, which this generation would like to use. Offering systems, which are not feature complete, would also invalidate a secondary aspect: If the population group with the least experience can effectively use an application, this application will be usable for any other, more experienced population group.

When dealing with systems for elderly citizens, the understanding of this cohort is the key to success. The rapidly ageing population has to be seen not only as a population group, but special features have to be defined according to the actual needs. Older people can understand e-government systems if they are designed for them as much as they are designed for younger people (Wandke, Sengpiel, & Sönksen, 2012).

Currently, the e-services which are offered by governments in the European Union as solution for this problem try to offer an easy and fast access to multiple layers of the government, but fail at the target group as several key systems are considered unusable by most citizens. E-government is gaining momentum around the world (Zhao, 2013), surveys by the UN (United Nations, 2010) show however, that there is a systematic divergence between systems offered and the actual adoption. The growing number of systems, the low demographic penetration and the ever-increasing complexity hinders the success of e-services and the digital divide between online and offline users will become deeper and deeper as systems become more complex with the increase of available functions.

The goal is therefore to find a way to map the needs of the elderly and create guidelines for the development of e-government systems. This challenge can be translated into research questions by building upon the two most important factors of perceived usability described by Maeda (Maeda, 2007):

1. "*The perceived complexity of the interface, which is created by the complexity of the task to be done and the design of the user interface*…" (Jung, 2002) The task itself should be considered a constant; therefore, the only way to decrease this factor is the improvement of the interface.
2. *"The previously acquired general experience of Internet applications…"* (Maeda, 2007), in our case Interactive e-services. This experience gives the user the edge when trying to solve a problem presented by the unfamiliar task and/or interface, thereby helping to overcome the perceived difficulties of the system.

The two areas are not only interchangeable, but complement each other to a certain degree. An easier to use interface does not need that previous knowledge, and the more knowledge the user has, the better he will be using a systems with bad usability (Maeda, 2007). The complement nature is however not ideal, no system can be designed so usable that no previous knowledge will be needed (Maeda, 2007). Therefore, the research goal is to minimize the necessary previous experience. The perception of usability by the user has to be considered an individual attribute, therefore designs geared towards extreme user-friendliness for a group of users is not always the best solution. This extreme can be observed in software or other tools designed with one and only one user group in mind. Computer mice designed to be perfectly formed for the right hand cannot be used by the left handed. Such designs can be considered optimal for their task, but will cause deep and unsolvable usability problems for users outside the target group.

E-government systems have to be built with the whole population in mind, and should therefore reject such design principles. Our goal is to find the optimal complement of guidelines for a usable design and models for improving the experience of the ageing population dealing with the systems, without creating applications where usability is improved by taking out features.

2. The Research question – Usability of e-Government Systems for the Elderly

Usability is the ease of use, and first of all the learnability of objects. This definition can be applied to any device or tool designed for human interaction. Usability is however also applicable to processes or workflows. The International Organisation for Standardisation (ISO) defines usability as follows: *“The extent to which a product can be used by specific users to archive specified goals with effectiveness, efficiency and satisfaction in a specified context of use…”* (ISO, 1998).

2.1 The Research Questions

The main question of our research is the usability of e-government services, in particularly for the elderly. As shown earlier, in chapter one, acceptance is the key indicator of usability of these systems.

Systems designed without concern about the users will result low usability and will be hard to use by the average user. User groups with experience below average, as elderly users, will reject these systems and continue to use offline procedures. As demographic projections indicate that this population group will become dominant in the European Union in the next decades, the problem stemming from the lack of acceptance of the systems by these users will rise to an unparalleled level. Governments will need to focus more intensely on the questions of acceptance of systems by elderly users. The need for cost reduction and requirements by the EU will make the diffusion of e-government unavoidable.

Therefore, our research poses the research questions around these problems:

- What are the measurable inhibitors for the elderly in e-government systems?
- What are the expectations of the elderly user for e-government systems?
- How can e-government systems be built in accordance to these expectations?
- Is it possible to build a standardised guideline, which results e-government system with higher acceptance, especially for elderly users?

All four research questions are centred on the interaction between an e-government system and the elderly users. According to Fisk et al. (Fisk, Rogers, Charness, Czaja, & Sharit, 2009) and Nielsen and Loranger (Nielsen & Loranger, 2006) the quality of this interaction is best measurable by the acceptability. This attribute is a definite unit of measurement for usability of the particular system. Fisk et al. and Nielsen et al. define five core factors, which determine the acceptability of a system, and thereby the usability of the device or application. In contrast, the ISO 9126-1 (ISO, 2001) standard for software quality describes usability as one of the core characteristics of software quality. In this research, we use the attributes defined by Fisk et al. and Nielsen and Loranger, as these describe interactive systems more precisely. However, according to Bertoa and Vallencillo (Bertoa & Vallencillo, 2010) both models are based on identical fundamental premises and are comparable. Below are the five attributes defined by Fisk and Nielsen and their analogous counterparts in the ISO model:

- *Learnability* is the capability of the software product to enable the user to learn how to use it. Learnability is of major concern in the design of complex applications, such as e-services. Learnability is also a characteristic defined by the ISO 9126-1.
- *Efficiency* implies that the product allows the user to achieve their intended objectives within a reasonable amount of time. This conforms to the core characteristic of *Understandability* featured in the ISO 9126-1.
- *Memorability* relates to how easy it is how to remember to use a device. This implies that the effort in relearning following periods of non-use should be minimal. This correlates to the attribute *User Compliance* in the ISO 9126-1.

- *Errors* can be interpreted as user actions that are performed or omitted and result in the user not accomplishing the desired goal. Feedback to the user concerning errors may or may not be signalled by the interface of the product. In any case, errors resulting from interacting with the product should be minimal and if they do occur, the user should be able to easily recover from them. This attribute is comparable to the ISO 9126-1 characteristic *Operability.*
- *Satisfaction* assesses the pleasantness of the experience the user has in interacting with the product. *Attractiveness* in the ISO 9126-1 is the comparable attribute.

Derived from the five characteristics is that the functionality and the usability of a product is interlinked. This has also been implied earlier and is one of the main issues, which makes the usability improvement process of e-government systems challenging. Removing features from a product will always improve usability, but as described by Nielsen (Nielsen J. , 1993) will at same time lower the perceived utility of the product. The example of senior phones is illustrative to this. The assessment of the functionally of the product and its link with the usability requirements is therefore a key aspect of usability research. The functionality has to be however view in respect to the intended users. The ISO 9126-1 (ISO, 2001) suggests, that the identification of the intended users is critical for the good usability of a system. As illustrated in the previous chapter the elderly have unique requirements when dealing with e-services. These will be used as basis for our research in order to answer the research questions. As we consider only fully featured systems, and will not imply usability improvement through removal of features, the results will be transferable to any population group with higher experience.

The ISO 9126 states that usability is a non-functional requirement of devices. This might not however be completely accurate for e-government systems. Usability has to be a functional requirement of such systems. E-services, which are functional, but not usable, have not fulfilled their core requirement. Nilesen and Loranger define this consequently: “If a device is difficult to use, people will not use it, they will leave” (Nielsen & Loranger, 2006). This seems to a very simple translation of the usability attributes described above. It is however accurate. Products, services or applications always have alternatives, which will be used, if the original is not acceptable. In case of e-government citizens will attempt to use new methods of interaction with the government only if it gives them clear and measurable advantages in communication, comfort or time required. If, however systems are not acceptable for the majority of the users, they will return to use traditional offline procedures. The methodology has to be built along these premises with measurable characteristics. The requirement for measurable attributes defining the usability was therefore translated into guiding principles based on the findings of Hartson (Hartson, 2003):

- *Realness* – Refers to whether a usability issue is a real problem or not. This category should be broadened to be able to categorise the occurring usability problems according to severity and amount (Nielsen J. , 1993). The impact of the problem can also be implemented as a category.
- *Validity* – is defined as the ratio of the number of real usability problems with respect to the total number of findings (classed as real and not real) (Jacko & Sears, 2003).
- *Thoroughness* – Identifies whether the utilised method finds all the real usability problems of the system. For the validity of this category a cross-examination of the system with multiple methods is required (Hartson, 2003).
- *Effectiveness* – A category defined by the validity and thoroughness of the method used. This is also in line with the ISO 9241-110 standard, which declares that effectiveness is the accuracy and completeness.
- *Consistency* – Is related to reliability and repeatability. Multiple usability inspection methods should produce “reasonably similar” results, when used on the same application (Molich, 2004).

Understanding the interaction of the user with the system is the key for the successful selection of a usability improvement methodology. As described earlier this can be accomplished by selecting a methodology, where the sterile metrics are augmented by user focused investigations. These are used to map the reaction of the users for specific processes and tasks. This helps to understand the context of the target group, in our case the specific problems of elderly users.

2.2 Importance of Usability Improvements when dealing with Elderly Users

As we defined earlier, Internet based G2C communication offers wide possibilities for all citizens, especially through its independence from location or office hours. For elderly users, these unique attributes of e-services can help to overcome mobility impairment or living in a remote location. Therefore, the rejection of these systems leads to severe problems in government projects, which define the one-stop-government as an objective. One-stop-government describes the concept of a single point of contact for the user for government services. This incorporates a single, aggregated information source for citizens. Online one-stop-government requires not only a fully networked government, but also a well-structured and understandable approach, meeting the perspectives and needs of the user. These systems would be the optimal tool for G2C communication for the elderly, requiring only a single point of contact, which would offer all information and thereby serve as a single governmental platform for any G2C communication. Current and past one-stop services, which are directed at the general population were often rejected with disappointing acceptance scores. The poor usability of these systems leads to a rejection by demographic groups with low experience. This has by definition a devastating effect for the acceptance by the elderly. A good example for this failure was the Bürgerkarte mentioned in the previous chapter.

As described by Maeda (Maeda, 2007), for elderly, this perceived sub-optimality of the usability can have two main reasons: The user is either inexperienced or strongly discouraged to use the application by internal aspects of the system. This is caused by psychological factors, attributable to previous experience or high-perceived complexity of the system (Maeda, 2007). These factors have to be identified to understand the context of the application in relation to the elderly users. Fortunately, the identification of both attributes are comparably straightforward to recognize when testing with the individual user. To solve the problem of low experience, users should be trained (Bruder C. , 2008) to the interface, thereby encouraging the usage and enhancing the users experience. Training is however not possible for systems where the interface is below a definite level (Bruder C. , 2008), which is at least considered acceptable by some users. Training users to the interface requires the maintenance of the mental image of the interface. This is however only possible for systems which are commonly used. This is achievable for mobile phones or other constantly used devices, but cannot be accomplished for e-government services, where the usage patterns are comparably infrequent. Therefore, the system itself has to be created with adequate consideration for usability. If the five core usability factors described earlier are not taken into account, or at least not with the targeted users in mind, the application will not be considered usable by the targeted users.

The solution for this problem is a usability-engineering based process, which should result in a system with a higher usability and acceptance. This process described by Rosson and Carroll (Rosson & Carroll, 2002) as "*user centred design*" is also suggested by the ISO 9241-210 (International Organization for Standardization, 2010).

The factors of the ageing population described earlier implicate several psychological factors, which have to be taken into consideration for e-government systems when inquiring with a focus on elderly users. One of the major factors is the cognitive changes and the learning ability at a higher age. Past analysis in this field can be quite clearly separated in two very distinct categories: analysis based on traditional learning (Merriam, Caffarella, & Baumgartner, 2007) and cognitive psychology with classic cognitive exercises. An example, which illustrates the second category, is the "Tower of Hanoi"[3], and similar tasks based on typical day-to-day scenarios. The analysis of the execution of such tasks by the individual is useful to answer the question about links between cognitive skills and performance of the individual when confronted with scenarios (Fleischmann, 2008). Classic cognitive exercises, like scenarios, are also useful to conduct tests in the field of usability research for Internet applications. These are strongly linked to day-to-day scenarios and are not meaningful when discussed alone. Even e-government services, which are not usually used on a day-to-day basis, can be considered in such context, as the interaction beyond the web-service is still known by the users and only the interface can be considered

[3] Mathematical puzzle consisting of 3 rods and a number of different disks. The goal of the puzzle is to move the stack of disks, which are initially ordered with the smallest disk on top from one of the rods to another by moving only one disk at a time and stacking disks only with decreasing diameter above each other, so that in the end all disks are again ordered by size.

unfamiliar. We use therefore this empirical method in our research to understand the cognitive capabilities of elderly users in relation to e-services.

An aspect, which stems from the incorporation of elderly users into a research of e-government systems, is the general acceptance of technology by the elderly. As shown earlier, elderly users are not unwilling to use the Internet (ARD/ZDF-Online-Studie, 2007), and multiple research papers (Bruder C. , 2008), (Grifoni, 2009) have shown, that elderly are interested in electronic devices. A survey conducted by the Ithaka research consulting in Hungary has come to similar conclusions. The number of "computer literate" elderly citizens has almost doubled in the last three years, with the numbers approaching 44%. The need for usable e-government systems is therefore present in both Germany and Hungary (Ithaka Consulting, 2009).

In our research we used the Technology Acceptance Model (TAM) by Davis (Davis, 1989) as a foundation to measure the general acceptance and to visualise the factors which govern the decision of the user. This model implicates that users are influenced by several factors when confronted by new technology. In case of the TAM 3 (Venkatesh & Bala, 2008) model there is a large number of different external and internal influences. This large number of factors incorporates a high number of variables to the model, some of which cannot be defined for e-services. Therefore we used the simpler and more reliable TAM described by Davis, with two basic attributes: perceived usefulness (U) and perceived ease-of-use (E).

Several methods are presented, which can assess several of these factors of a system in relation to the user. These offer an array of evaluation types, from inspection through scenario based inquiry to actual testing on systems, and present a comprehensive overview of the user-system interface. The selection of an utilisable evaluation technique required a comprehensive look at the advantages and usefulness of the different procedures:

- *Think-aloud protocol* – A method favoured by Fisk et al. (Fisk, Rogers, Charness, Czaja, & Sharit, 2009) to gather data in usability testing. It has been developed based on the Würzburg School of Psychology at IBM for interface user testing by Lewis (Lewis C. H., 1982). Think-aloud protocols involve participants thinking aloud as they are performing a set of specified tasks. Users are asked to say whatever they are looking at, thinking, doing and feeling, as they accomplish the task. This enables the observer to see first-hand the process of task completion. Test sessions might be video and/or audio recorded, in addition to capturing the screen interactions. The results are relatively close to the real user experience and can therefore be used to effectively model the actual usage. The main disadvantage is however the lab based user assessment, which might offer a sterile and unnatural environment for the tested users. Nonetheless, the think-aloud protocol is one of the few methods, which can offer a deep insight into the actual perceived experience of the user. Therefore we used this method in the research to gather information about the perception of the system by the user.

- *Focus groups* – It is a form of qualitative research, in which a group of people are asked about their perceptions, options, beliefs and attitudes toward a system or concept. Questions are asked in an interactive group setting where participants are free to talk with other group members. This method can be used to better understand the motivation of the users and their concerns about the application. The output is however purely subjective, no objective data is gathered by focus groups. In addition, the unusual environment and the group setting can result in inaccurate and unrepresentative data (Coolican, 2009). This procedure was discarded as not suited for pinpointing the genuine user expectation, as the members in a focus group pose an unpredictable influence on each other.
- *Semi-structured interviews* – These can be used as an assessment of individual experience and expectations, without remote effects from other users. It can provide a very detailed and deep understanding of a system, but requires comparably few participants for relatively representative results. Standardisation of questions and fixed responses enable the gathering of data, which can be compared and analysed. A major disadvantage of interviews is however their time requirement, as a one-to-one approach is extremely time consuming, when compared to other evaluation methods. This procedure was used in the first phase of the research to collect information about the expectations.
- *Cognitive walkthrough* – It is a heuristic method used to identify usability issues in software, focusing on how easy it is for an inexperienced user to accomplish a task

with the actual application. The method is rooted in the concept that users typically prefer to learn a system by using it to accomplish tasks, rather than from manuals or helps. It offers a tool to refine the requirements of the system and does not necessitate a fully functioning application. It can use prototypes which lack some elements. It is however not usable to quantify user satisfaction. In addition, experts and designers are not the "traditional" users, and as such behave differently. This method provided a framework for the phases two and three of our research.

- *Pluralistic walkthrough* – A heuristic expert-analysis method, used to identify usability issues by reviewing a paper prototype of the system by experts and usability engineers. The method is centred on a scenario-based review of the paper prototype, thereby resolving usability issues of the system. This enables a relatively fast and efficient usability analysis, which offers an overview of the major problems of the application. The main disadvantage of the method is comparable to other expert based techniques, experts are not the typical users, specialists cannot find some usability problems. This procedure enabled the selection of an actual system, which has been used for the usability testing in phases two and three.

The selection of the effective methodology was the groundwork for the comprehensive evaluation in our research. As described earlier, representative results for interactive systems need scenario based testing beyond heuristic methods. The previously described underlying complexity of e-government applications also determines the selection of the correct methodology. As described by Ziefle and Bay (Ziefle & Bay, 2003), the very low experience of elderly users may offer a challenge in the usability improvement process, therefore the low experience has to be taken into account in research projects. Our methodology therefore includes an analysis of the experience of the actual users in every phase of our research.

3. The Research Hypothesis

We build the workflow of our research on the research questions defined in the previous chapter. These can be simplified into one inquiry: *"Is it possible to create a generalised solution to usability problems of e-government systems for the elderly?"*

This question leads to the formation of a hypothesis, that through solving the age-specific usability issue, systems can be successfully offered for the elderly. We also postulate that such systems will result in satisfactory levels of user acceptance. Furthermore we hope to achieve results which can be generalised in order to build a framework for a comprehensive solution of the problem.

3.1 The Formation of the Hypothesis

As stated in the previous chapters, the main idea behind our research came from the usability aspects of e-service systems. This focus was built upon 3 assumptions in mind:

- Usability can be influenced by a better user interface in case of e-government applications. This has been successfully demonstrated by projects at the University of Potsdam in 2008, where the acceptance of different service platforms for the general population was improved by a usability engineering process (Molnar & Nguyen, 2008). This project focused on the general usability issues without a emphasising on age-specific problems.

  This improvement can also be observed for appliances or interactive tools for elderly users. This has been shown by research projects in the field of cognitive ergonomy, such as Bruder (Bruder C. , 2008). Bruder and others focused strongly on the elderly as a specific user-group and used a wide range of appliances ranging from mobile phones to car navigation systems. E-government systems tested by other projects were only one-way non-interactive systems built around citizen portals (Molnar & Nguyen, 2008), or focused solely on e-forms (Sayago, Sloan, &

Blat, 2011). Highly developed interactive e-government systems however were not tested until recently. User interface analyses by Holzinger et al. (Holzinger, Searle, Kleinberger, Seffah, & Javahery, 2008) and Becker (Becker, 2005) conclude that extensive work is needed to enhance the quality of such systems for elderly people.

- A further factor, which has an impact on age-specific issues, is the importance of acceptance. This concern is becoming increasingly significant as demographic challenges present new problems, some of which can only be met by new, innovative G2C communication. It can be assumed that this can be solved by user-centred design with a focus on the elderly. The problems relating to the ageing population mentioned earlier were the catalyst for the initial decision to concentrate on this user group. This focus is mainly built on the demographic changes in the European Union, which means that any future service issued for the whole of the population has to be usable in particular by this growing cohort.

- A further issue is, that systems created for this population group should be usable and acceptable by the complete population, thereby creating applications, which are aligned with the external requirements of governments. This statement is based on the idea that systems designed for the population group with the statistically lowest experience will also be usable by any set of citizens with higher average experience. This also poses the question about the age-specific usability versus the general problems of systems. This issue is regarded as one of the key components of the hypothesis. The age-specific problems have to be balanced against the feature completeness of the system in order not to cause a system where the above idea about a generalised solution is no longer valid.

- A concept which governs our research is the subject of Internet penetration in relation to advancements in e-government systems. Currently, high-profile systems are commissioned to give the user a “one-stop-government” and an easy access to everyday governmental services. This approach, which is considered to be superior to any offline system by giving more comfort and easy-of-use to the citizens, is reliant on acceptance.

3.2 The configuration of the data acquisition

The main objective of a study into an array of systems, where the general concept is the inclusion of the population as whole, is to find the optimal complement of guidelines for a usable design. A secondary goal is the construction of models for improving the experience and the knowledge of the ageing population dealing with the systems. Both objectives have to be accomplished without creating systems that have an improved usability at the cost of removing features.

These objectives were derived from the assumptions declared in the previous section, and are resolved with a framework built upon the theoretical groundwork. The synergy from applied psychology and applied computer science combines information gathered from interviews and behavioural studies with a technological analysis and interface redesign.

This methodology was developed in accordance with findings of analogous research projects for other products for the elderly (Bruder C. , 2008), (Sayago, Sloan, & Blat, 2011). As shown earlier, the most comprehensive methodology for a usability assessment of products is the integration of inspection methods and interface tests. This is also supported by multiple authors (Nielsen J. , 1993), (Fisk, Rogers, Charness, Czaja, & Sharit, 2009) as the only method which results in a comprehensive overview of the system. Inspection methods used should include heuristics, cognitive walkthroughs and action analyses. Developers often use these methods to inspect and assess the general usability of a service before implementation. This however does not include the future users into the process, and might result invalid assumptions about the actual requirements. Seale et al. (Seale, McCreadie, Turner-Smith, & Tinker, 2002) postulate three main reasons for involving older users in the development process:

- First, application requirements can be set out clearly and further problems can be avoided.
- Secondly, stereotypes of technology minimising elderly handicaps are reduced.
- Thirdly, cost-effective design for all products can be developed with early participation. Studies by Czaja and Lee (Czaja & Lee, 2007) also show that inclusion

of older users strongly improves the perceived usefulness of the resulting application.

These theories and concepts have been developed into an integrated usability improvement method for our research. This serves as a foundation for an iterative usability improvement process created along the hypothesis that heuristics in combination with dedicated user input will not only result an improved system as postulated by multiple authors, but also lead to a solution which can be used to create a generalised guideline. The process can be seen in figure 3.

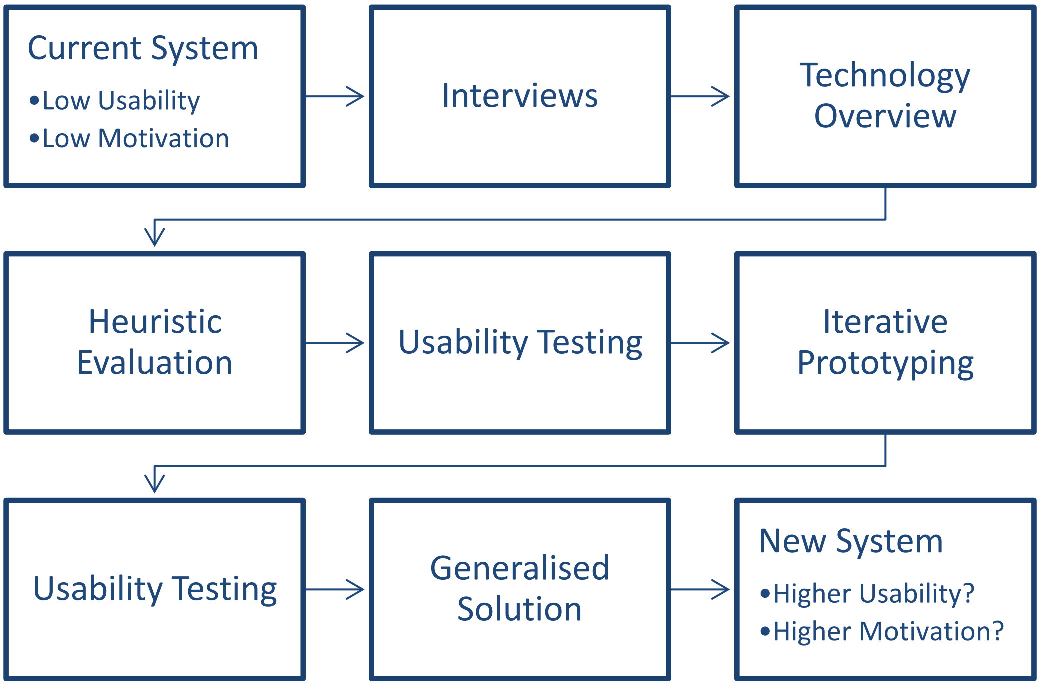


Fig. 3. - Integrated usability improvement process – broad schematics of the research

As described above, and seen on the schematics of the research, the initial premise was that currently offered e-government systems not only offer low usability, but also have measurable problems motivating users, especially the elderly. This idea was further developed into the theory that systems should work for the user, and be aligned with the tasks, which are considered significant by the target group. The general framework of the project was therefore built upon three phases, which include multiple proven principles of human factors approaches. These enable a repeatable method, which results in a more usable and therefore more accepted interface, thereby also increasing the user motivation to use the applications. The decision to create a research process based on three phases was based on the principles of design by Fisk et al. (Fisk, Rogers, Charness, Czaja, & Sharit, 2009). The phases were built around the concepts of task analysis, empirical measurement and iterative design. An integrated design approach as described by Fisk et al. is however unpractical when dealing with e-government systems, as the complexity of these e-services makes such parallel evaluation extremely time and cost consuming. Therefore, a hybrid process model was developed based on the findings of Fisk et al., but with a workflow built along the ISO 9241-210. A detailed description of the research steps can be found in the next chapter.

3.2.1 Interviews and Technology Overview

This step forms the basis of the research, and was designed to give an overview about the needs and the general problems of the target group. It also includes a technology overview, which is basically a generalised expert analysis on currently offered e-government systems in Germany, Hungary and on European level.

This step is also important as an assessment of the future direction of the technology, which might govern the development of systems in the next decade. It thereby helped to understand the future problems and issues, which the target group may encounter. The interviews have been created on a purely abstract basis, with no actual e-government system serving as a model. This decision was essential, as any functional system might have influenced the test group with actual and real problems, which consequently would have resulted in falsified data based on a single system. The interviews were therefore based on a system, which made an analogous approach in both countries possible and would give the elderly several advantages through improved service quality and efficiency. The idea behind the service was designed along pre-interviews, which gave information about the factors, which would be seen as an improvement by the target group.

Among other ideas, a theoretical “retiree-card” application was specified, which was centred on the concept of typical German university student IDs. The scenario for the research step was constructed around the request procedure for this card and based on the maturity levels of e-government systems declared by the European Commission. This approach guaranteed that the users were confronted step-by-step with more and more electronic components and were not overwhelmed by new concepts, which might have created refusal or other sudden and uncontrolled changes in the attitude of the test candidates. Following the data gathering, the technology available was analysed to create an overview of the available systems, which could serve as a platform for further research.

3.2.2 Heuristic Evaluation and Usability Testing

The data gathered from the first step enabled us to identify a system for further tests, which served as a test platform. This system had to maintain the following criteria:

- The system had to be user-centred, with a service offered for citizens in a G2C configuration.
- The context had to be relevant for elderly citizens and independent from language, nationality, social, educational or cultural background to make a cross-cultural study possible.
- The quality of the system had to be of a level not to disrupt the tests through incompatibilities, system crashes or other software problems.
- The complexity of the selected service had to be of a certain level. Only systems with medium or high complexity can adequately simulate the usability problems for a successful later generalisation.
- The openness of the application was requirement in order to use it as a platform for the iterative system development.

These criteria led to the selection of a system, which was in line with the efforts of the European Union to create a single Pan-European Identification Solution (Atkins, 2004), and concluded in the selection of the new German eID System, which was introduced in Q3 2010 in Germany. This system might be followed by comparable systems in other EU member states (Naumann & Hogben, 2009). From a technological standpoint, it is not only highly sophisticated but also complies with the stringent German data privacy laws, which can be considered the most demanding of any EU member state. The data collection in this phase of the research was centred on the hypothesis that the system had to be usable for a wide array of citizen to government (C2G) interaction.

The heuristic evaluation of the designated system was followed by the selection of actual scenarios, which would fit in the user context. These scenarios were chosen according to the data collected in the first phase, which indicate that electronic interaction is generally favoured when dealing with specific services. The scenarios were only used as a background for the user-tests as the main aspect remains the user analysis of the new ID-Card with the “AusweisApp” software client. The main goal of this research step was the actual user-test, which was created in accordance with quantitative user-testing methods and focused on laboratory based testing based on procedures by Krug (Krug, 2010). The scenario-based model was divided into three steps for each candidate, with the pre-test computer knowledge and the post-test experience measured. The main sequence was used to analyse the usability of the client through the actual test, based on the scenarios above. The tests were conducted with every test candidate once with each scenario, thereby measuring the learnability of the interface. The data gathered from these tests in Germany and Hungary enabled us to build new interface prototypes, which served as showcases for a generalisation of usability rules and practices for e-government systems when dealing with the ageing population.

3.2.3 Development of a solution according to the gathered data

The final step in the research was the creation of a generalised framework and the proof of concept with an iterative development cycle. This allowed us to create prototypes according to the actual needs of the users uncovered through the framework. This concept was based on the success of similar methods in other projects (Bailey, 2005), (Tan, Liu, Bishu, Muralidhar, & Meyery 2001). This development method was based on a spiral software development established by Boehm (Boehm, 1986), which is combining both design and prototyping phases, thereby bonding top-down and bottom-up methods. A further refinement of this cycle with usability centred design can be created when using this method along common usability models, thereby improving not only the functional aspects of the software, but also the usability for the target user group.

A method consisting of not only prototyping, but also small-scale user testing of the actual prototypes further enhanced this improvement effect. This made possible a deeper understanding of the needs of the target group thereby allowing further usability tailoring of the client to the measured usability requirements. The development stage of the research was followed by a summative evaluation of the final prototype. These final tests were conducted in accordance with the previously used analysis and evaluation methods, thereby delivering not only verifiable data, but also enabling a direct comparison with the original client. This comparison also permitted us to isolate key weaknesses in the original application, and thereby support the general usability framework.

4. **Empirical Studies**

This chapter includes a detailed sequence of the research based on the theoretical approach introduced in the last section. It contains the empirical studies which have been developed in accordance with the usability improvement process introduced. The actual quantitative goals have been derived from this process for an effective planning of the phases:

- **Phase 1** – The analysis of the general acceptance of e-government service by the elderly.
- **Phase 2** – Understanding the context of the system and the connection between acceptance and the system attributes.
- **Phase 3** – Proof of concept of generalised usability centred software development methods for e-government applications.

These phases were selected according to the principles of user-centred design declared by Fisk et al. (Fisk, Rogers, Charness, Czaja, & Sharit, 2009) and are built on the concepts formulated earlier. We assume that this method not only illustrates the main problems of current e-government development, but also postulates an applicable and proven guideline.

The first question, which has to be answered, is the basic question of e-government in relation to elderly users: *"What are the main attributes of an acceptable application for elderly users? What functions do they need and how will they interact with the application?"*

These questions served as basis for the phase model of the research, and have been asked in similar form by several other researchers, especially in connection with human-computer interaction (Fisk, Rogers, Charness, Czaja, & Sharit, 2009). In case of e-forms in e-government systems, preliminary results by Righi et al. (Righi, Sayago, & Blat, 2011) show that the use of e-government can be promoted to older people. They profit from the functionality according to Righi et al., but the research did not answer the questions about details of interaction.

Therefore, in our empirical studies we also focused on this, beyond the inquiry about the required functionality. A method has been selected, which is not only user driven, but can also be adapted for specific requirements in the development process. The basic model set out by ISO 9241-210 standard (International Organization for Standardization, 2010) conforms to these constraints, as it describes a cycle which can adapted to specific software development goals. Furthermore it is based on the idea of a user driven improvement process. The selection of the ISO 9241-210 as the basis for the process enabled therefore the use and further modification of the human-centred design activities. According to the standard not only the user and the task itself define the context of the system and its usability attributes, but also the physical environment and the purpose of the application. It must be declared, however, that the ISO standard sees the user as the most important aspect when designing an application (Deutsche Akkreditierungsstelle, 2010). This factor was seen as the most defining attribute in the selection of a method. Merging this into the hypotheses describing the user behaviour and the technology acceptance of the elderly enabled to extend the model to better represent our target group and to specify the application environment. This allowed the creation of a hybrid model based on the ISO 9241-210 activity diagram, describing the tasks of the research. These lead not only to a better understanding of the requirements, but also to a general solution for e-government systems.

This flow-model can be seen in figure 4 with the inner cycle representing the standard ISO 9241-210 and the outer, superimposed cycle, the modified sequence that was used as the basis for our research.

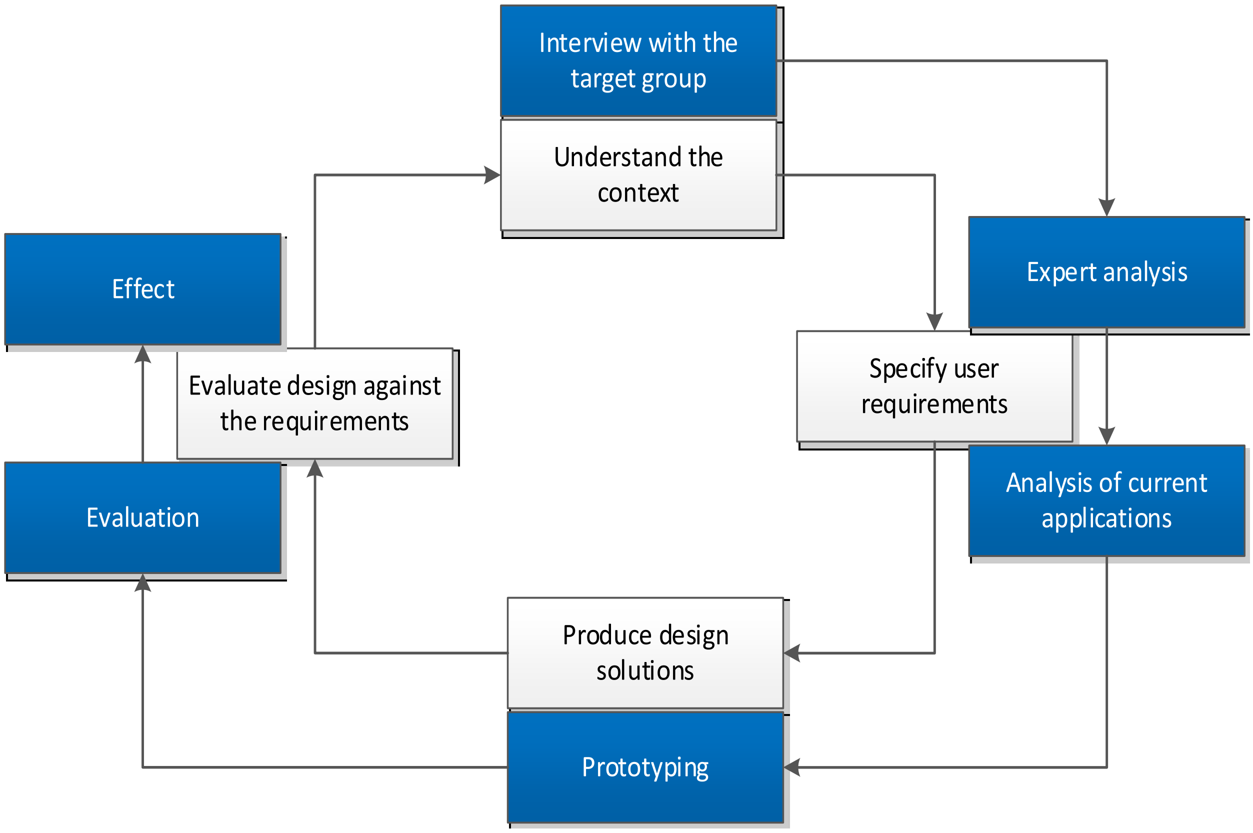


Fig. 4. - Flow-model of the research based on the ISO 9241-210 software development diagram, Source: (International Organization for Standardization, 2010)

4.1 Phase 1 – Understand the context – Interviews

It can be observed in the ISO 9241-210 based flow-model, that the analysis of a particular usability problem should be started with the users as the principal focus. This provides a way of understanding of the actual intentions of the users and creates a solid foundation for further research. This also enables a clear focus on the actual user requirements in the second step of the flow-model. We used a basic method in this phase, which created qualitative data about the requirements of targeted users. Probable qualitative field study techniques, which were considered, were case studies, focus groups and interviews. The investigation for optimal methods resulted in the selection of a deep interview technique, which was based mostly on fixed responses. These were created on the premises of similar studies of user and task analysis by Hackos and Redish (Hackos & Redish, 1998). Methods of comparable user context analysis were built on deep interviews for a complete assessment, combined with scenario based role plays which offered a deeper understanding of the user context, giving information about the main goals of the users and their problems with the theoretical system. According to Rosson and Carroll (Rosson & Carroll, 2002), interviews also enable an easy to learn environment for the user, where important general information can be discovered directly and effectively in a short time. The success of the scenario based method is however extremely dependent on the validity of the scenarios, therefore it is imperative to select scenarios, which are relevant to the targeted users. Results from interviews help to understand not only the general problems that the users face, but also the tacit needs that the most users, who are described by Prensky (Prensky, 2001) as "digital-immigrants" cannot formulate. Deep interviews enable a connection between statements like "*it is unusable*" and the real cause for these assertions. The interviews also offer a possibility to assess the global context of the concept in relation to the targeted users.

This conclusion also supports the concept of scenario based methods for the assessment of the user's point of view. This is in particular useful, if the users are not professionals and/or have a very low previous experience with the systems at hand. In case of our research, the interviews were accomplished with a target group consisting of participants with a low level of previous computer and Internet experience. This meant that additionally the language used in the scenarios had to be chosen in accordance with this issue to avoid problems from missing experience as well as from the unfamiliarity with the technical terms. According to Wixon et al. (Wixon, Holzblatt, & Knox, 1990) user needs can only be understood through methods, where the users can connect the new systems to a task or an older system, which they are acquainted with. The lack of such connection to a known factor will not only give incorrect results, but can also lead to a massive rejection. A new system coupled with a set of new tasks can prove to be a challenge. If even the objective, the concept and the terms of the system are new to the participants, the rate of refusal will be even higher and might cause invalid results. In accordance with this theory, in order to obtain a higher cooperation from the participants, we did offer scenarios with a familiar concept.

4.1.1 The Task

This phase of the research was built around a concept, which was regarded as familiar by the target group. A retiree-card was conceived as a model for these scenarios, which included the application and delivery procedure for the new document from the first step to the actual hand out. The retiree-card as a scenario is consistent with the theory of Wixon et al. (Wixon, Holzblatt, & Knox, 1990) and offers the participants a scenario, of which the core concept can be considered to be familiar for them. To assess the different reactions to the different e-government approaches, the scenario was scaled along the maturity levels of the European commission introduced earlier:

- *Level 0* – Offline System with no complex electronic components and only traditional face-to-face or postal communication. The scenario was designed to mimic the problems of such procedures. Information was not completely available before actually visiting the fictional office as the limited content given to them about the retiree-card was only on a pamphlet representing the basic information. The paper-only-communication was represented by not issuing any forms or documents before the process.
- *Level 1* – One-way electronic system with passive information available. These applications offer the user information about the process, opening times, data requirements or other relevant content. These pieces of information were given to the user through a paper guide, which represented the non-interactive website of the process. Apart from the guide, the scenario was identical to the first level.
- *Level 2* – One-way electronic system with passive information and online forms available. Such systems enable the users to retrieve information about the process and also to fill out the relevant forms before printing them out. The forms however still have to be taken or sent to the designated office manually. This maturity level has been simulated by an additional possibility offered in the scenario, where the form has been shown to the participants before starting the process. A similar paper-guide as before was offered. According to Lines et al. (Lines, Ikechi, & Hone,

2007) elderly users of e-services benefit strongly from the previously handed out forms which can be filled out before the procedure.

- *Level 3* – Two-way fully integrated e-government system. These applications represent the pinnacle of e-government, and provide the user with the opportunity to receive all information and fill out all forms electronically. It also allows the user to send them by electronic means to the government. Personal interaction is however necessary for the hand out of the final document. This level also incorporates by definition electronic support. The possibility of personal contact is still offered at this level. This has been simulated by an identical structure as previously, but with a possibility for questions. The information necessary to fill out the application form was given to the participants on a sheet with a short help for each question.
- *Level 4* – Fully electronic service without personal contact. The process was described to be fully electronic, without the need to personally visit any governmental office at all. This level was simulated along identical premises as level 3, with the subtraction of the option of personal contact. The possibility of a system without direct personal contact was analysed through this level.

Additionally, socio-demographic data was gathered. This was needed to assess the context and the preceding experience of the test candidates before the actual scenarios. Beyond gathering information about the education, gender and age, several questions enabled the assessment of prior knowledge and experience with computer and Internet habits. The relevant questionnaires in German and Hungarian can be found in Appendix A. The experience with computers was asked through a guided question: Answers enabled 4 different replies from "I have never used a computer" to "I use a computer almost every day, and can accomplish complex tasks with it." Any interaction beyond email composition and Internet surfing was considered complex at this point.

A detailed classification was done in the second phase of the research. The participants were also asked about the actual ownership of a computer at their accommodation. Any Internet capable device was considered a computer. Additionally, the experience with actual of e-government systems was also inquired. This was accomplished by a yes/no question. These questions enabled an analysis of the premises of the participants and their actual experience with comparable systems. Information about this aspect is essential, as these can have a substantial effect on acceptance and attitude when incorporating more and more electronic components into the scenario, thereby removing more and more traditional face-to-face interactions with the government.

4.1.2 The Sample

The sample for the interviews were 70 citizens from Germany and Hungary between the age of 62 and 79. 45 participants were gathered from the federal states of Berlin and Brandenburg. The Proportion of genders was equal in the sample. The tests in Hungary have been conducted with 25 test-participants form the metropolitan region of Budapest, with comparable parameters as in Germany. The summative median age was 69.5 years.

4.1.3 Independent and Dependent Variables

The maturity levels introduced before are the independent variables in this phase of the research. These were the basis for the interviews and were used as variables. The additional variables were persistent and enabled a comparison of the results in relation of the maturity levels. The comparability of the acceptance of the fictional e-government system was based on dependent variables consisting of the three standard questions described below and based on the After Scenario Questionnaire (ASQ) (Lewis, 1991). This implicates that the measured acceptance is the sum of the

- perceived utility (ASQ_{pu}) of the fictional e-government system of the relevant maturity,
- the perceived time-gain (ASQ_{pt}) through the implemented electronic components of the service,
- the perceived easiness (ASQ_{pe}) of the fictional task at the relevant maturity level.

The supplemental and the socio-demographic questions were used as control variables. The statistical results were analysed to gather information about the users themselves. This enabled the first step in understanding the context of e-government systems for the elderly.

4.1.4 The Procedure

The acceptance of the levels was measured with a standardised questionnaire, consisting of three standard questions of the ASQ method, shown in the previous section, based on the research by J.R. Lewis (Lewis, 1991) at IBM. We have modified the method for the use with the scenarios, as the basic framework places its emphasis on multiple choice grading. This was not deemed effective in case of scenarios based on multiple maturity levels. A simple yes or no answer gave sufficient data for the assessment of the basic user needs at this level. Later in the research, multiple level grading was introduced as it provided a finer distribution for the results.

In addition, the standard ASQ questions were accompanied with supplementary queries, needed for a deeper understanding of attributes unique to e-government. This procedure can be seen in figure 5.

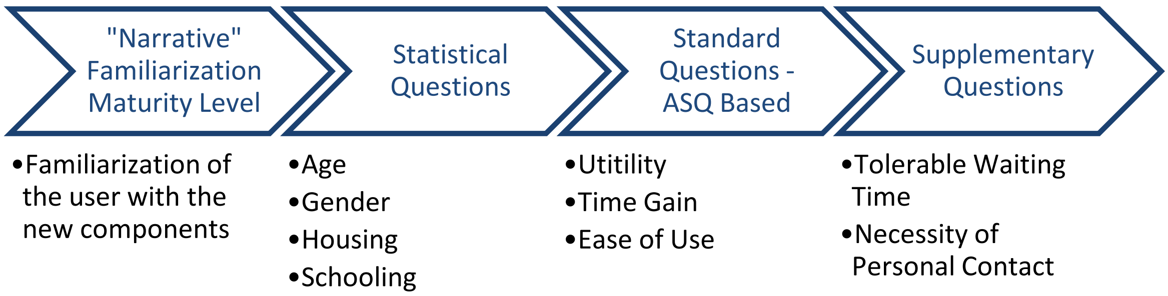


Fig. 5. - Model of the interview procedure

The three main aspects of the ISO 9241 mentioned and described earlier – efficiency, utility and effectiveness – link the standard with the ASQ, leading to comprehensive and reproducible results. Utility in the ASQ is consistent with utility in the ISO 9241. Ease of use is consistent with effectiveness and time gain is consistent with efficiency. The final ISO 9241-210 was not yet available at the time of the design of the model in late 2009; therefore, some details of the final standard could not be employed. This is also the main reason for the use of the older ISO 9241 and ISO 13407 standards at this stage of our research.

Before the questionnaire, the participants were familiarised with the scenario by a short verbal introduction consisting of the narrative overview. This was followed by the relevant questionnaire. This was repeated for each maturity level with the relevant narrative part and then the questionnaire. The standard questions were asked for each maturity level. The supplementary questions were asked at the relevant maturity levels, thereby creating a deeper understanding of the user experience of highly advanced systems. This was followed by the socio-demographic questions, consisting of age, gender, housing, schooling and previous experience with e-services. The interviews were based on a standardised questionnaire consisting of multiple questions arranged into two distinct groups:

- Standard questions, based on the ASQ method:
    - Are the new electronic components helpful?

      This question enabled the assessment of the perceived utility of the particular scenario level. It is consistent with the standard utility question of the ASQ model.

    - Is the process faster with the new system?

      The perceived time gain was measured through this inquiry, thereby enabling a comparison between the maturity levels of the model. The idea behind this question is also established on the standard ASQ, with modifications for a multi-tier scenario based method.

    - Would you use this system?

      This question has been adapted fundamentally to the inquiry about the perceived ease-of-use of the task. This was based on the research of other research groups, who had very promising results with this modification (Ramayah & Ignatius, 2002). The intention of future usage can be regarded to be a strong indicator of an easy-to-use system. The intention to use an application is also one of the key aspects of the TAM (Davis, 1989), and strongly affects the actual system usage.

- Supplementary questions:
  - The tolerable maximum waiting time for an offline transaction has been measured for a deeper understanding of problems of traditional G2C interactions. The possible answers for this multiple-choice question have been given from zero to “more than 60 minutes” in intervals of ten minutes.
  - The perception of the received information has been queried for an additional visualisation of the differences between the incremental maturity levels. The results helped to comprehend the definite requirements for delivered information in case of different maturity levels.
  - For the most advanced level of the scenario a final question has been designed. On this maturity level, the simulated system is an online-only process, which enabled G2C communication without the possibility of any face-to-face interaction. This purely electronic channel opens the question about the actual need for personal communication in G2C transactions. This was therefore measured through a question about the acceptability. This question also measured whether this is required by elderly users.

The answered questionnaires were analysed and evaluated in order to visualise the theoretical acceptance of e-services for elderly users. The theoretical acceptance was calculated from the three ASQ questions by assigning one or zero for positive or negative answers. The theoretical system was considered as accepted if the three scores were positive for the user.

4.1.5 Evaluation of the results

The socio-demographic data was used to create an overview about the external factors affecting the sample group. As seen in table 1, the education level of the participants was approximately consistent between the countries. The strong overrepresentation of highly educated participants and the underrepresentation of test applicants with primary education as their highest schooling is an occurrence, which is common in usability research concerning high-tech devices. Higher educated individuals are statistically not only more affluent, but also predominantly more interested in new opportunities and technologies. This is also one of the basic factors of the Technology Adoption Lifecycle model (Rogers, 1962). According to Rogers, education and technology adoption have a statistically significant correlation.

Table 1 – Highest education level of the test participants in Hungary and Germany in comparison to the general education levels in both countries, 2009. Sources: (Központi Statisztikai Hivatal, 2012) (Statistisches Bundesamt, 2011)

N_{Sum}=70	Hungary		Germany	
	$N_{Hungary}$=25	General Population (All cohorts)	$N_{Germany}$=45	General Population (All cohorts)
University or College	48.00%	17.50%	33.00%	13.50%
Secondary Education	44.00%	54.90%	58.00%	57.82%
Primary Education or no Education	8.00%	27.60%	9.00%	28.68%

Table 2 – Previous experience level of the test candidates in Hungary and Germany, 2009

N_{Sum}=70	Hungary $N_{Hungary}$=25	Germany $N_{Germany}$=45
High	0.00%	4.00%
Middle	28.00%	38.00%
Low or None	72.00%	58.00%

The assessment of the previous experience with computers and the Internet coupled with the data gathered from the previous question allowed a broader understanding of the sample and the requirements of the cohort in relation to e-government. The deficiency of previous knowledge observable in table 2 visualises the "digital divide" described by Mehra (Mehra, 2004). This effect is the division of the population through the experience with electronic devices. This is even an issue in a sample where higher education is overrepresented. Table 3 shows this effect with a cross-sectional analysis of the young and older elderly.

Table 3 – Ownership of a computer by the test applicants in Hungary and Germany, 2009

N_{Sum}=70	Hungary $N_{Hungary}$=25	Germany $N_{Germany}$=45
Young elderly (Under the age of 70) N_y=37	54.00%	71.00%
Older elderly (Over the age of 70) N_o=33	17.00%	23.00%

As seen table 3, there is some difference between the countries, with a higher percentage of computer ownership in Germany, but this factor does not stem from the "digital divide". It is caused by the prices of computers compared to the average income of elderly in the different countries.[4] The conclusion of this data is that although older elderly would profit more from the Internet and electronic services, these cohorts have very little access to these. This can also be shown through the correlation between the age and computer ownership in Germany and Hungary. There is a weak $(r_{Germany} = -0.28, N_{Germany} = 45)$ through moderate $(r_{Hungary} = -0.38, N_{Hungary} = 25)$ negative correlation between the variables. The correlation between age and computer ownership is at $\propto = 5\%$ is significant in both countries: $(t(0.95) = 1.68,\ t_{crit\_D} = 1.91\ ,\ t(0.95) = 1.71,\ t_{crit\_H} = 1.97)$. This can also be supported by data gathered later in the second phase of the research with more granular and detailed methods. This however also supports the results, that elderly are starting to be marginalized as digital services become more and more prevalent, and that the "digital divide" will become an issue which has to be dealt with.

4.1.5.1 Level 0 – The Offline System

The scenario was conducted with the attributes mentioned above. The process presented to the participants was a fully offline application without any electronic components.

[4] The average pension in Germany is 788 EUR/month (Deutsche Rentenversicherung, 2009), the average pension in Hungary is 88 000 HUF/month, which is 350 EUR/month (Országos Nyugdíjbiztosítísási Főigazgatóság, 2009). The costs for an entry level computer are identical in both countries, at about 400 EUR.

Communication was simulated as being reliant on traditional channels, as phone, postal service or brochures. This, combined with the simulated issues described earlier resulted in a very strong rejection of the process. Common statements by the users included "*cumbersome*", "*too complicated*" or "*I have better things to do*.

The acceptance measured through the ASQ showed that 68 out of 70 participants were generally dissatisfied with the simulated procedure: The score for the intended usage was zero for all but two users. This was however not unanticipated, as such systems produced similar results in other papers (Molnar & Nguyen, 2008). In contrast, the data acquired from the question about the tolerable waiting time produced surprising results. Waiting time was defined as time spent waiting in a sitting position at a governmental office while being required to actively monitor a display, which showed the number of the next client.

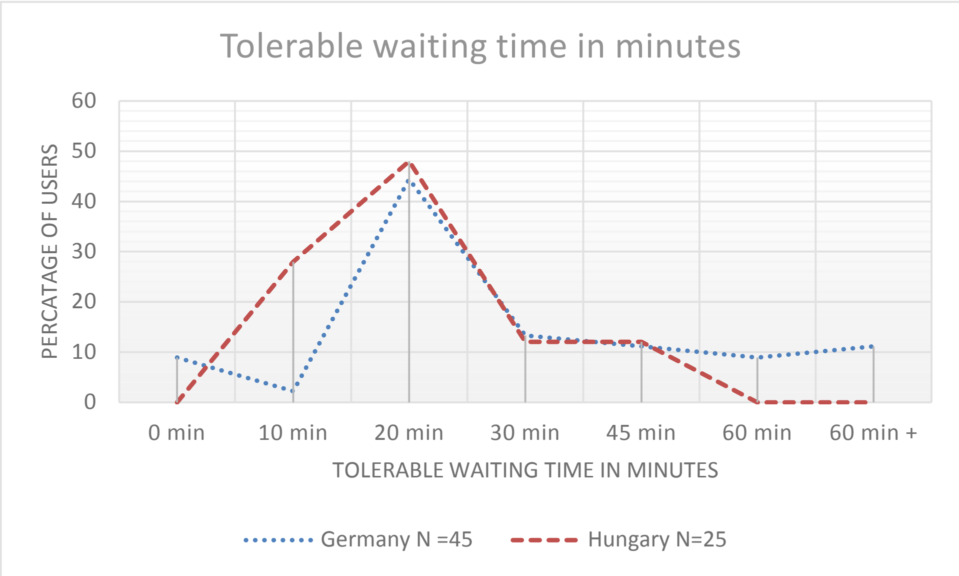


Fig. 6 – Tolerable maximum waiting time by the test candidates in Hungary and Germany, 2009

Figure 6 shows the distribution of tolerable waiting time. The maximum of this curve is at 20 minutes in both countries. Data acquired in Germany has somewhat more extremes, which is not surprising, as patience in Germany seem to be more highly regarded culturally as in Hungary. The peak of the distribution at 20 minutes illustrates the problems with offline systems. Users are not eager to wait long times and the waiting time has an impact on the acceptance of the system. Despite of the cultural differences and the somewhat different distribution between the countries, the peak of the distribution is at 20 minutes in both counties. The different flattening of the curve at the extremes has however no significant effect on the usability of the application.

4.1.5.2 Level 1 – One-way Electronic System with Passive Information

The second level of the scenario was based on a system which integrated a simulated website displaying content about the process of acquiring the retiree-card. This enabled the participants to gather information before starting the procedure. This was simulated by a paper model which contained information about the retiree-card, and explained, which agencies are involved and which forms or declarations will be required. This maturity level has been extremely common and widely available in both countries since 2004-2005. Examples include governmental portals (www.brandenburg.de, www.magyarorszag.hu) and further websites offered by governmental agencies throughout the vertical structure of both governments. Therefore we believe, that most citizens, who have used e-government systems before, have most likely used applications of this maturity level.

The influence of this could also be observed in the results: The intended usage and perceived utility scores of this maturity level reflect the previous contact with e-services. 61 users ($N = 70$) answered all three ASQ questions positively, supporting the theory that this maturity level has an extremely high acceptance. Only four participants answered the question about the perceived utility with no (two in Germany, two in Hungary). The "is the process faster?" was answered by 67 participants positively. 65 of the 70 users tested answered the "would you use it" question with yes.

4.1.5.3 Level 2 – One-way Electronic System with Online Forms

The system for this level was expanded with the potential of electronic forms, which could be downloaded and filled out electronically, before printing them out. This unidirectional, interactive approach enables a more efficient and faster access to the government, as citizens are offered the opportunity to fill out forms before getting in personal contact with the government. Presently, numerous systems offer this type of access, predominantly for processes requiring multiple complex forms, which might even refer to each other. Tax declaration and other finance based processes started to employ systems of this maturity in the last three to four years throughout the European Union. The use of these is however not compulsory for private persons in most countries. On the other hand, businesses have to declare their taxes electronically in Hungary since 2007 (Magyaroszag.hu, 2006). This has caused unforeseen consequences, as in some cases citizens are forced to use a system, which was designed for tax consultants. According to law (2005. CLXIII.), small businesses with a single employee have to declare their taxes electronically, requiring them to use a level 3 system, which was not designed for this user group (Futó, 2007).

We have also observed effects of this in our scenarios. Younger elderly in Hungary have significantly higher theoretical acceptance threshold than younger elderly in Germany. The combined results measured for this level are still acceptable with a summative acceptance of 59% $(N = 70)$. The results are however not so remarkable, if the acquired data is dispersed into younger and older elderly. Our results are also supported by the findings of Lines et al. (Lines, Ikechi, & Hone, 2007). According to their research, training elderly in e-form completion is possible and observably reduces and/or eliminates the usability problems when using online services. Lines et al. only observed this for younger elderly, older elderly did not benefit significantly from training.

Table 4 – Acceptance of a Level 2 e-government system by the elderly user in Hungary and Germany, 2009

$N_{Sum}=70$	Hungary $N_{Hungary}=25$	Germany $N_{Germany}=45$
Younger elderly (Under the age of 70)	77.00%	58.00%
Older elderly (Over the age of 70)	48.00%	52.00%
Correlation between age and acceptance (Pearson's product moment correlation coefficient)	-0.18	-0.12

The differences in table 4 can be easily explained by two distinct influences: The higher acceptance in Hungary for younger cohorts is caused by the higher experience of these users through the training effect described above. Additionally, elderly under 70 were still actively part of the workforce in the mid-1990s when the digital revolution caused the mass-diffusion of computers in offices in almost every field.

It can be stated however, that this maturity level is still accepted by the majority of the elderly users, and such systems are considered useful, efficient and effective. The diffusion according to the theory of technology diffusion (Rogers, 1962) has reached a level, which suggests that such systems can become generally accepted with better usability. The diffusion level of this maturity level is also described by multiple researchers, who agree that this level might become common for e-government systems in the next decade (Wind & Kröger, 2006), (Kohlhofer, 2009).

4.1.5.4 Level 3/4 – Two-way Electronic System with and without Personal Contact

These systems can be considered the pinnacle of e-government system design. They provide the fastest, most efficient communication between the citizens and the government. As maturity levels 3 and 4 only differ in the process control and the possibility of personal contact, we have considered it unnecessary to test the user on both maturity levels. The process control aspect is only significant for the back office procedures between agencies, the users are unaffected by these. Therefore the two maturity levels could be combined for our research. The difference between them was measured by the question about the need of personal services.

Every step in the process is accomplished on the Internet, without the need for direct personal contact between the citizen and the government. Information, forms and additional content is available online, and presented for electronic communication without the need for postal or personal delivery of forms. In addition, processes beyond the G2C interface are also fully electronic. Agencies are offered an integrated process control and communication system, which enables the efficient processing of documents even between multiple governmental offices. The single offline component of the process is the delivery of documents, in our scenario the retiree cards are delivered to the elderly by postal service.

This concept is theoretically effective and efficient; the interviews implicated however, that level 4 systems might not be considered acceptable by certain cohorts. The fully electronic process caused several negative statements such as "*too complex*", "*not clear how to use*" and "*where can I get help*". These statements evidently show that these applications might need more thought when offered to users with little or no previous experience with e-services.

Table 5 – Acceptance of a Level 3/4 e-government system by elderly users in Hungary and Germany, 2009

N_{Sum}=70	Hungary $N_{Hungary}$=25	Germany $N_{Germany}$=45
Younger elderly (Under the age of 70)	54.00%	54.00%
Older elderly (Over the age of 70)	8.00%	29.00%
Correlation between age and acceptance (Pearson's product moment correlation coefficient)	-0.43	-0.24

The differences between the cohorts might originate from the factors described at level 2, the age at retirement has an even stronger impact on the results. This is also supported by the correlation between age and acceptance. Test participants under the age of 70 have shown some acceptance for a system without direct human contact, and about half of this cohort seems to value efficiency over direct personal communication. As for the older participants, this factor becomes extremely important and this cohort strongly rejected the fully electronic approach. The differences between the two countries shown in table 5 can be attributed to the stronger centralisation in Hungary. The shorter distances to citizen offices and other governmental agencies also contribute to the lower acceptability of absence of direct personal communication. Communication based help was seen important by fewer than 40% of the younger participants ($N_y = 37$); older candidates ($N_o = 33$) have however objected this to be crucial, with over 80% of the answers rejecting the electronic approach.

4.1.5.5 Conclusion of the scenarios

The results from the different levels analysed in accordance with the maturity model of the European Commission seems to support the hypothesis that there is a threshold in the acceptance of e-government and probably other types of e-services for different cohorts.

This threshold can be observed in figure 7. The results demonstrate a distinct drop in acceptance for the older elderly at level 2. Simulated approaches rejected by at least one-half of the participants should be considered as not accepted and such applications will probably suffer from general acceptance problems if offered to the population in their current design.

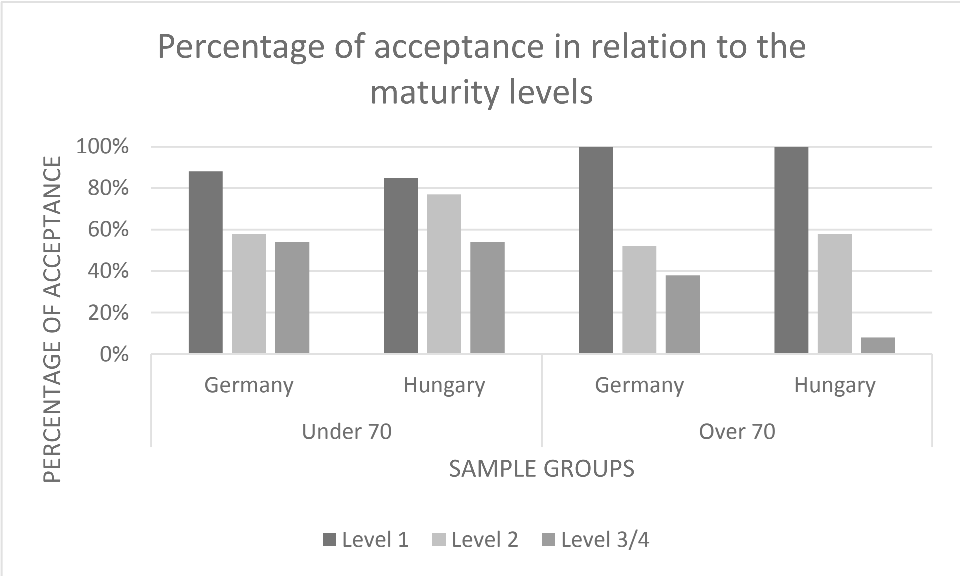


Fig. 7 – Comparison of the acceptance of the scenario levels by the elderly users in Hungary and Germany, 2009

The results support the theory that younger elderly approve e-government systems which offer electronic components up to level 2. Level 3/4 systems were somewhat more controversial, but even such systems were acceptable for about 50% of the younger test participants in both countries. Older elderly might however not accept present systems beyond level 1. Interactive systems seem to encourage a fear of new technology and disapproval in about 50% of the tested participants. Further research was therefore concentrated on this issue, as these cohorts could profit the most from new offerings. Level 3 and 4 electronic applications could offer the most for this age group, as it would enable easy and comfortable G2C communication and could support the efficiency of ambient assisted living through a communication platform with the government.

4.2 Specify User Requirements for Elderly Users

In line with the ISO 9241-210, the next step of our research was the specification of the user requirements for e-government systems for elderly users. A general conclusion of the first phase was that electronic systems are not disapproved by the elderly, on the contrary, applications up to level 2 are considered useful. The detailed analysis of these systems however needed an approach, which could be generalised, and therefore be repeatable with any appropriate application. This was archived with a concept built upon two different sub-phases: The user requirements can be first specified in a white-room environment with a heuristic evaluation, creating a framework for further tests (Isbister & Schaffer, 2008). The approach by Isbister and Schaffer was fine-tuned and used in the selection of the optimal system for the later tests. We used this method, as the broad spectrum of e-government systems and the vastly different goals of the diverse services give such an immense array of dissimilar applications that without a heuristic evaluation, the selection of the optimal system would be almost impossible. The framework of the evaluation was built upon an array of conclusions from the first phase, which the designated system should comply with. These provisions enabled the development of a decision-routine, which served as a selection tool for the ideal system for the necessary tests.

This approach ensured the selection of an optimal system, which was coherent with the requirements uncovered in the first phase. In the second part of this research phase, the system selected with the help of the decision-routine was used as the basis for the user evaluation. The needs of the users could thereby be verified on an actual system, and used in the later, third phase for the optimisation and the creation of the generalised solution for the elderly. The user tests were built on a scenario based approach, with different life-like scenarios constructed around the real system. The scenarios were selected in accordance with the results of the interviews.

4.2.1 The Heuristic Evaluation

The main issue of this chapter was the selection of a suitable system, which was achieved through the process described earlier. The key requirement for successful tests was not only the compliance of the selected system with the maturity levels defined by the European Commission for e-government systems, but also that the system had to offer a service approximately at level 3 – a two-way-electronic system. This was imperative as it would comply with the critical threshold of acceptance uncovered in the first phase. As we have shown earlier, this was the most advanced level which can be considered to be acceptable by younger elderly users. Systems have to comply with the requirements of these users, as they will become the majority of the elderly in the future.

Furthermore, this maturity level is also considered to be the aspired maturity level for most systems. Level 4 systems without any personal interaction can be implemented for several services, but are difficult to experiment with, and might also increase rejection as suggested by the results of the first phase. This factor eliminated several possibilities for test systems, which are often used for similar research projects. Citizen portals and other largely static services only offer one-way communication between the citizens and the government, and were therefore not applicable for this project.

A technological factor we took into consideration, was the issue of modifiability of the chosen system. This aspect was one of the key questions, as the system selected in this phase of the research had to be used later in the prototyping and evaluation. The requirement of translation into Hungarian (or German) also contributed to this question. Only systems with open architecture were considered for the tests, or services where the GUI (Graphical User Interface) could be separated from the application without significant loss of key functional components. The second option was based on the idea of retro-engineering the GUI for the requirements. This constraint meant that only recently developed systems could be taken into consideration.

A further technological aspect was the question of system reliability and responsiveness. Gould and Lewis (Gould & Lewis, 1985) found by testing appliances that functional problems could not only interrupt the tests, but also falsify the results by creating dissatisfaction and disapproval by the users. Their objection is also true for e-services, as the rejection of the system would not originate from the usability, but from the instability of the system.

Additionally, the tasks, which the tests are based on, have to be considered significant by the targeted users. This is a critical characteristic, as with any other exercise, a process which is considered unimportant by the user leads to results which cannot be considered significant. For an optimal user requirement analysis, the chosen service has to fit into the lifestyle and the activities of the targeted user group. In case of our research, processes not considered useful for the test group, consisting of elderly and retired, were eliminated early in the analysis. For an optimal selection process, these requirements were translated into an UML (Unified Modelling Language) decision process (Briand, Labiche, & Sauve, 2006), where the factors could be analysed and thereby lead to a reproducible and well documented result.

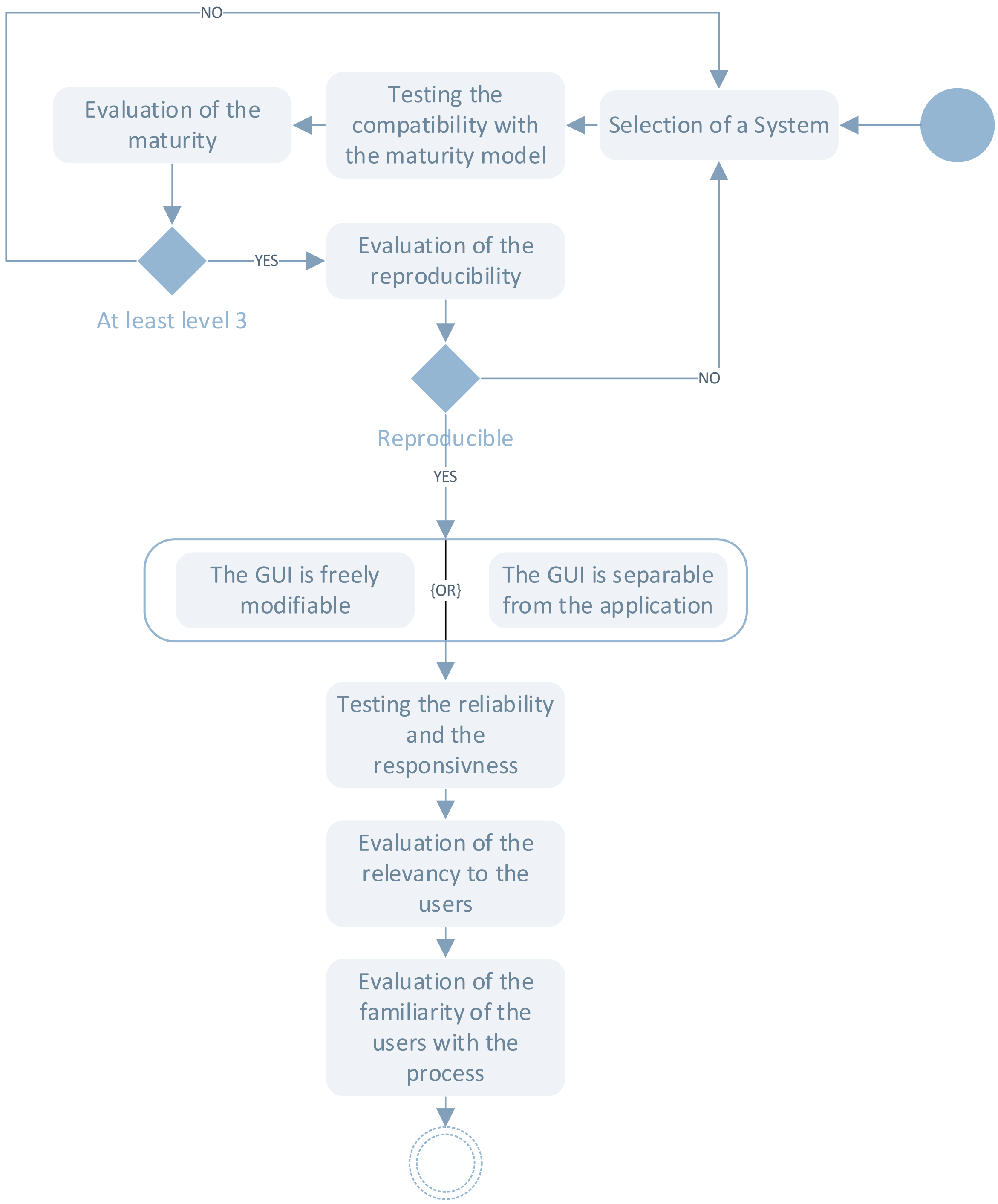


Fig. 8 – UML - Decision process for the heuristic evaluation (based on Briand, Labiche, & Sauvé, 2006)

The evaluation was accomplished along the UML process, and enabled a detailed inquiry into considered systems. These premises enabled the inclusion of several systems in the heuristic evaluation. This allowed the selection of a suitable system through structured elimination along the set criteria. The elimination steps were developed and selected along the criteria described earlier.

The first applications taken into consideration were governmental portals; these systems have a GUI (Graphical User Interface), which is easily customizable by design. Governmental portals are level 1 or 2 systems, which are considered to be acceptable by elderly users. The usability of the citizen portal www.brandenburg.de was analysed by a team from the University of Potsdam in 2008 (Molnar & Nguyen, 2008), with several users over the age 65 interviewed. The results indicate that government portals are viewed as favourable and considered useful by approximately 65% of the cohort. These systems are however not complex enough to serve as the foundation for a deep understanding and are not suitable examples which a standardised solution can be built on. Integrated G2C communication platforms were also considered, in particular, the Hungarian “Ügyfélkapu”[5]. Although, these system were considered ideal for usability testing, they proved to be unusable, as the GUI is an integral and inseparable part of the application. This makes a translation of the system into German unfeasible. Electronic tax or pension declaration systems were also rejected on the objection of low importance and familiarity to the users.

[5] The system is an integrated service platform for G2C applications.

System	Advantage	Disadvantage	Suitability for the Research
Citizen Portals – e.g. Brandenburg.de	• Viewed favourably by the elderly	• Level 2 maturity • GUI is integral part of the system	Not suitable, system maturity does not reach level 3
Integrated G2C communication platforms – e.g. Ügyfélkapu	• Level 3 maturity • Reliable and Responsive	• GUI not open • GUI not separable from the application	Not suitable, user interface is integral part of the application, therefore not modifiable for testing
Electronic Tax or Pension declaration systems – e.g. Elster	• Level 3 maturity • Reliable and Responsive	• Users unfamiliar with the concepts • Importance for the users limited • GUI not separable from the process	Not suitable, users can not identify themselves with the functionality
Electronic component of official documents – e.g. AusweisApp[6]	• Level 3 maturity • GUI separable • Important for the users	• Concept unknown to users, scenario based approach needed • First versions were unstable	Suitable, as the system does satisfy all attributes set out by the heuristic evaluation

Fig. 9 – Summary of the heuristic evaluation

The system analysed and considered suitable to be used for the user tests, was the electronic G2C component of the then new German ID-Card. This client was not only in line with the maturity model but also presented a flexible setup, enabling the possibility for a comparison of multiple maturities on the very same system. It is a client-server application,

[6] The „AusweisApp“ was called the „Bürgerclient“ before November 2010. The first tests have been made with these pre-release versions, which were however feature and GUI-complete in October 2010. The first release version (4.1) was identical with this test version.

which can be reengineered, rebuilt and modified according to the needs of the research, without requiring any modification of the server application.

The versions available in late 2010 could be considered reliable and responsive in a controlled test environment. Several problems of the client have been uncovered recentlywhich might compromise the security of the transactions. These problems have not interfered with the tests in this step of the research, as the security or the client-server communication of the system was not the target of the analysis.

The communication of the software had also been restricted, as any storage of personal data would have caused unforeseen legal consequences, and would have complicated the research enormously through privacy related procedures. The objective of the research did not need the communication and data storage components of the system, as these aspects are not considered significant when dealing with the usability of a GUI by the end-user. A further advantage of the processes behind the system was the simulation of everyday tasks, which the users were familiar with. This enabled the disconnection of the process from the GUI, and provided a better view on usability, without interference from problems with an unfamiliar process. This lead to tasks engineered which were considered to be useful for the tested cohorts in both countries. The selection of the “AusweisApp” of the new German ID-Card also enabled us to use a system, which was in development and introduction in 2010. This permitted some possibilities for modifications and gave us an opportunity to use some of the infrastructure of the developers at the Fraunhofer FOKUS for the initial tests.

In addition to being a suitable system from the technological and user point of view, the prototype AusweisApp made available by the Fraunhofer FOKUS also contained the scenarios needed for the proposed approach. Later modification of the test setup could be easily redesigned through the standardised components and the standard languages in the development. The AusweisApp was built as a Java™ client, but could be used with any scenario created in a normal HTML (Hypertext Markup Language) or XML (Extended Markup Language) environment, thereby providing a modifiable framework for the tests.

This potential was used in the Hungarian tests and in the next step of the research to create optimisations without modifications on the core components. The comparability and reproducibility of the test runs was thereby highly enhanced.

Fig. 10 – Schematics of the Test Set-Up

The functionality of the "AusweisApp" in relation to the HTML page the users were first confronted with can be seen on figure 10. This shows the basic steps of the GUI and the configuration of the client based testing environment. This also illustrates the steps, which were implemented into the "AusweisApp" process to enable a task independent operation of the client. The user can launch the "AusweisApp" client by clicking on a button on a standard HTML web page, the client then takes over and offers an interface for the secure identification of the user. This is achieved through mutual identification by a certificate of the service provider and a certificate and PIN (Personal Identification Number) of the user. This is followed by the data transmission and a confirmation on the HTML page.

4.2.2 The User Tests

The next step after the selection of the tests system was the set-up of the infrastructure. Two dummy applications were selected from the possibilities offered by the Fraunhofer FOKUS. Both applications were extended by creating a scenario around them. This was proposed to offer a life-like environment for the users and to make the scenarios feel more genuine. The Fraunhofer FOKUS institute built scenarios according to earlier acceptance tests, therefore the scenarios themselves were not analysed in our research, we focused only on the client environment. In addition, the client was considered to be the primary GUI for the G2C interaction.

4.2.2.1 The Sample

Analogous to the interviews in the first phase of the research, the test sample has been selected with matching premises for the user tests. The number of candidates has been set to 75, 45 from Germany and 30 from Hungary. The sample size is high enough for significant results. Calculations in G*Power 3.13[7] show that for these independent samples from Germany and Hungary a sample size of 23 for both groups is sufficient.

As in the interviews, the participants were recruited from Berlin and Potsdam, and the metropolitan area of Budapest. The test candidates were in the cohorts between 60 and 78 ($N = 75, M = 68.27, SD = 6.35$). Gender was evenly distributed (47% female, 53% male).

In addition, a control group consisting of 20 students of the Humboldt-University of Berlin was recruited to be able to evaluate the results in relation to an optimised user group. This group consisted of participants between the age of 20 and 34 ($N = 20\ , M = 27.95, SD = 3.95$). Gender was evenly distributed with a ratio of 50/50.

4.2.2.2 The Task

The first scenario was an online banking service, where the user could open a bank account without any personal contact with the bank, solely through authentication by an electronic ID-Card. This process was embedded into a scenario consisting of the opening of a bank account at a hypothetical large bank, which offered better condition for savings than the current bank of the user: "*You have read that this large bank offers better conditions for your savings, but in order to obtain these, the account has to be opened through the Internet*…" This statement started the scenario, where the users were confronted with the system described earlier.

The second scenario was based on an electronic video download service offering titles which needed age verification. This process was embedded into a fictional scenario, where

[7] http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3

the user found an old film not obtainable otherwise. *"You have a found a long-time favourite film of yours on this film download service. The film does however need age-verification, which can be done with your new ID-Card…"*

In both scenarios the users were confronted with the client seen in figure 11 (German) and figure 12 (Hungarian).

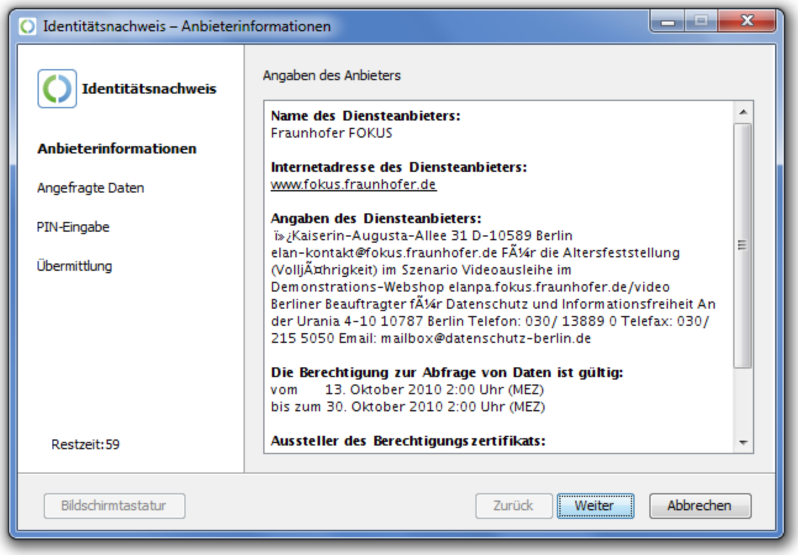


Fig. 11 – "AusweisApp" Version 4.1 – Certificate Information Screen (Source: Application created by the Bundesministerium der Innern/Fraunhofer FOKUS, 2010)

The attributes described previously made it possible for the full test environment to be translated into Hungarian. The AusweisApp, available only in German, was re-engineered and rebuilt in a VisualC++.Net™ environment to enable the identical tests in both countries without the language barrier. It must be noted however that difference which resulted from the lexicographic deviances between German and Hungarian could not be fully remedied. Some elements of the interface had to be designed to the conditions set by the lexicography of Hungarian. The size of some buttons or labels could not be set to be identical as in the German version, this would have resulted a smaller font size, which in turn would have influenced the comparability of the GUIs. It can be declared however that the two versions are identical in function and nearly identical in the GUI with no impact of the language on the scenarios.

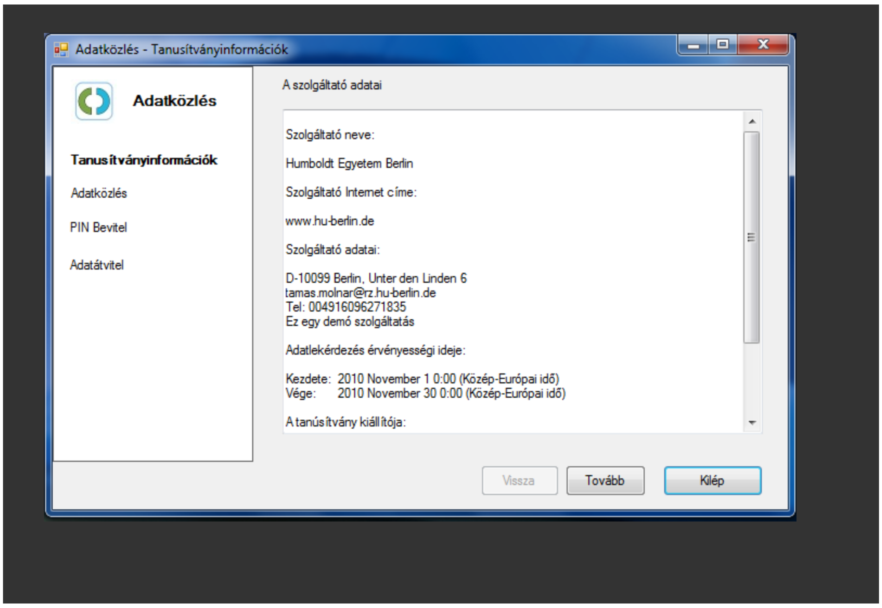


Fig. 12 – "AusweisApp" Hungarian Version – Certificate Information Screen (Source: Based on the "AusweisApp" created by the Bundesministerium der Innern/Fraunhofer FOKUS, 2010)

This method coupled with identical scenarios in both countries enabled the direct comparison of the results.

4.2.2.3 Independent and Dependent Variables

The main goal of the tests was to gather data about the usability of systems in an everyday environment with users of the selected cohort. The tests were therefore based not only the scenarios, but also backed-up by a methodology designed for the assessment of the user experience and proven by similar research projects. To completely understand the users' previous familiarity with comparable computer systems, a complex analysis of the experience on an individual level was needed (Maeda, 2007). This was based on the CLS (Computer Literacy Scale) developed at the chair of engineering-psychology at the Humboldt-University Berlin by Sengpiel and Dittberger (Sengpiel & Dittberner, 2008). This model can assess the computer literacy of the user through a series of questions, creating a matrix of typical tasks on a computer. The questions of the model had to be somewhat modified to be in accordance with the e-government environment. Questions regarding previous e-government system usage were added, expanding the model thereby to assess existing experience to a full spectrum of interactive Internet based applications. Only part A of the model was used to assess the users experience with computers.

In order to assess the Internet habits, the ownership of Internet capable devices, computer, tablet or smartphones was asked. Additionally questions were set up to gather information about the computer experience and usage of the candidates. This was used to create a broader picture about the experience of the cohort.

The actual tests with the AusweisApp were conducted according to the data gathered in the first phase, and they made up the main part of the user tests. They were performed with a thinking-aloud method to document the accomplishment of the tasks. The mouse movements on the screen were captured by screen-capture technology. A direct audio or video feed of the tests was not captured, as it would have highly complicated the set-up by creating the need for strict privacy agreements. In addition, fear from misuse of privacy relevant data would have made the recruitment process extremely hard in this case. This deficit was however compensated by the application of the RSME (Rating Scale Mental Effort) developed by Zijstra (Zijstra, 1993). The RSME enabled the scoring of the mental effort on a one-dimensional scale of 0 to 150, where 0 is no mental effort at all and 150 is an unbearable mental effort. The users were asked to indicate the effort they needed to complete the tasks. This enabled, combined with the observations through the thinking-aloud method enabled an overview over the user perception of the scenario. During the scenarios, the number of external advices or assists needed was recorded and categorised. We measured the seriousness of the encountered problems by analysing their impact (I) on the successful competition of the task (Nielsen & Loranger, 2006).

The impact of the problems was classed on a scale of one to three, with minor issues as one and problems with critical outcome for the success of the task with three.

- Minor impact would likely derail the user from the task, but should only lead to a temporary setback, resulting in a longer completion time needed for the particular scenario. The impact for the process is minimal. The user can recover from the error without external help.
- Medium impact problems would very likely lead to serious ramifications, resulting in the need for support and strongly affecting the perceived ease-of-use of the system.

Such problems are probably recoverable after external help. Without external help these errors might be unrecoverable if the experience of the user is low.

- Critical impact problems would likely cause the user to give up the task and cause the complete rejection of the system. These problems can be considered mission critical. These errors are unrecoverable without external help. Even with external help, some users will abort the task if the mental effort reaches a critical level.

Additional crucial data was acquired by the after scenario questionnaire (ASQ) which was adapted for these user tests. The three questions, which the ASQ is based on, were used according to the results from the first phase, thereby contributing to the comparability of the results. The ASQ itself was based on the work of Lewis J.R. (Lewis, 1991):

- *“Please mark the perceived difficulty of the application.”*
- *“Please mark the perceived usefulness of the application.”*
- “I’m satisfied with the time I have needed to complete the application.”

The answers were provided on a scale of 1 to 5 from strongly disagree (1) to strongly agree (5).

4.2.2.4 The Procedure

The test environment has been set up for three different iterations of the user evaluation. A pilot-test with a small number of candidates was essential to verify not only the test equipment, but also the scenarios. This was conducted in October 2010 with the same premises as the later tests. The usability laboratories of the Fraunhofer FOKUS were used for the pilot-tests and for the test runs in Germany with the planned number of test participants. The eight volunteers for the pilot-test were recruited from the best possible candidates (20-30 year old university students). This cohort has theoretically the highest previous experience with similar systems and would easily accomplish the set out tasks, thereby verifying the measurability and utility of the scenarios. This step enabled to modify the tests before the essential evaluation. The pilot-tests were conducted with a think-aloud method and with the standardised questionnaires based on the methodology presented. The finalised set-up was used to gather reference data for the tests. This was accomplished with a second test-run with a sample of five additional volunteers from the same cohort. This enabled to set up the ideal course of usage by gathering information on the behaviour of the ideal users. This data was later used as a basis for the tests to be able to create a comparison between older and younger participants and to visualise the complexity of the usability deficits of the system.

4.2.3 Evaluation of the Results - Tests with the cohort in Germany and Hungary

The tests have been conducted in parallel in both countries to assess the general usability of such systems on a European level. This unique set-up enabled a broader understanding of the special requirements when dealing with elderly users. The tests groups were recruited from community college courses specially offered for the elderly, and from basic computer courses offered by senior leisure centres in Berlin, Potsdam and in metropolitan area of Budapest. Test candidates recruited with this process possessed basic computer skills. Users without any previous experience with computer would not use e-government applications, as long as any other alternative is provided by the state. Additionally, the tests with computer illiterate users might not provide usable results as these participants would first need to learn the basic concepts of computers before any tests can be done. According to Scherer (Scherer, 2008), this causes the users to focus on the problems of the physical user interface, which is leading to results which are not measurable in case of a software usability analysis.

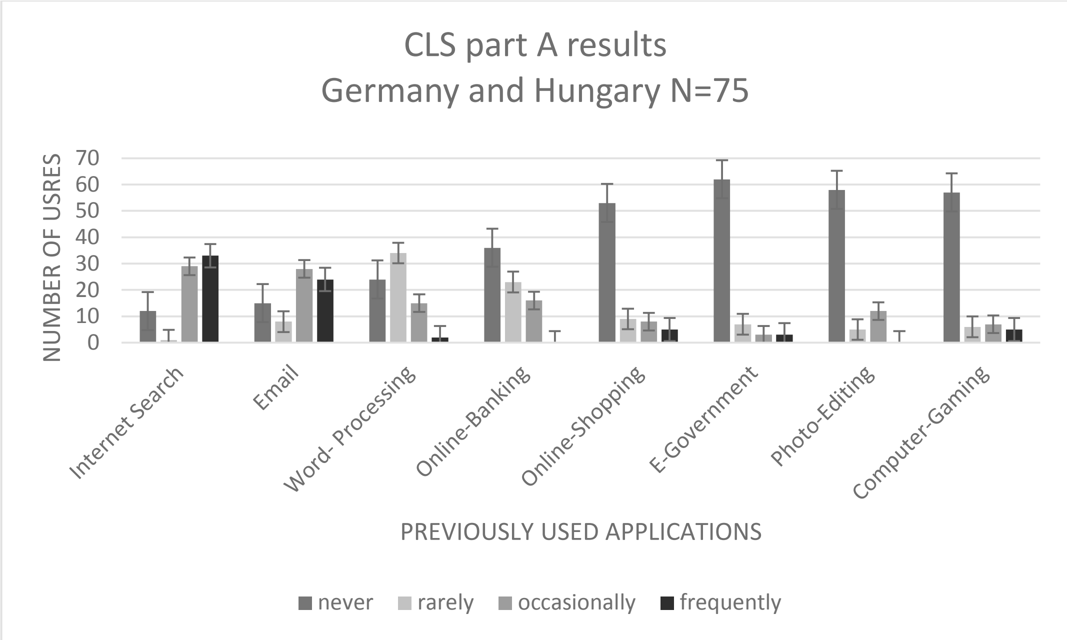


Fig. 13 – Results from the Part A of the CLS (Computer Literacy Scale) with 75 participants (Source: based on (Sengpiel & Dittberner, 2008))

We used the Part A of the CLS as an analysis of the experience of the users. The data was collected through a matrix where the users could mark the appropriate system usage (See Appendix A). This was accomplished individually in the usability lab of the Fraunhofer FOKUS in Berlin and the lab of the Corvinus University in Budapest. This minimised the internal interference of the data, thereby increasing the validity of the outcome. The results from the CLS (figure 13) show that, as we estimated, elderly users have a basic experience when dealing with computers. Almost 70% use Internet search, email and word processors frequently or occasionally. Numerous elderly use online banking rarely or occasionally. Interactive applications, as online shopping or e-government are used only by a small percentage of the elderly and seem to cause more problems.

These applications require more complex interactions from the user, and they are therefore more dependant on the user having experience with similar systems. Results from the previous phase of our research are also consistent with the results from the CLS. It shows that highly interactive systems are perceived as challenging by the elderly and have a lower acceptance. The usage of Internet banking is somewhat in contradiction with this statement. Although highly interactive, such services have been used by 55% of the users. This might be explained by two factors. Offline banking is becoming more and more costly and is seen is increasingly cumbersome when compared with the online alternative. Additionally most online banking transactions are very straightforward interactions, as most processes only involve the use of the interface to check accounts and pay bills occasionally. The interactions are however standardised processes. These are also mostly analogous between different banks in the same country, further increasing the factor of a known interface (Mahn, 2006).

The CLS was followed by the scenario based user tests, which supplied information about the problems the users were facing with actual systems. The direct usability of the system was measured by the problems encountered. The encountered problems, which were categorised earlier, were counted while testing the users. When encountering a problem, users were encouraged to complete the scenario after assistance from the test leader.

In order to visualise the number of encountered problems from the three categories in relation to the issues of elderly users, we also tested the system with experienced users. For this purpose 20 students from Humboldt-University were recruited. The addition of this test group also made it possible to visualise the cross-sectional differences between younger and older users of the same system. This group of users was termed control group.

Table 6 – Mean number of problems encountered by the users – Scenario 1 (Source: based on (Nielsen & Loranger, 2006))

N=95	Minor (1)	Medium (2)	Critical (3)
Germany $N_{Germany}$=45	3.21	1.57	1.82
Hungary $N_{Hungary}$=30	3.8	1.63	1.8
Control Group $N_{Control}$=20	1.75	1.11	0.2

As seen in table 6, there is no correlation (See appendix C for detail) between the mean number of encountered problems and the nationality. The minimal discrepancy can be explained with the somewhat lower CLS scores in Hungary. The total number of encountered problems can be seen in figure 14.

The main issue, which causes the high number of medium and critical problems seems to be the structure of the application. This has also been observed by other papers (Wandke & Jörn, 1999), and is described as one the main reasons leading to critical problems of the users. This is caused by the low previous experience, which has a high impact on the use of an interface (Maeda, 2007).

The large number of critical problems, which have often led to an abort of the process, is however alarming and visualises the deficits of the GUI. Only the comparably highly experienced users of the control group accomplished the scenario mostly without critical problems encountered. It is desirable to create an interface, where critical problems are not encountered, as these can cause severe distress for the users and lead to negative motivation through failures and disappointment.

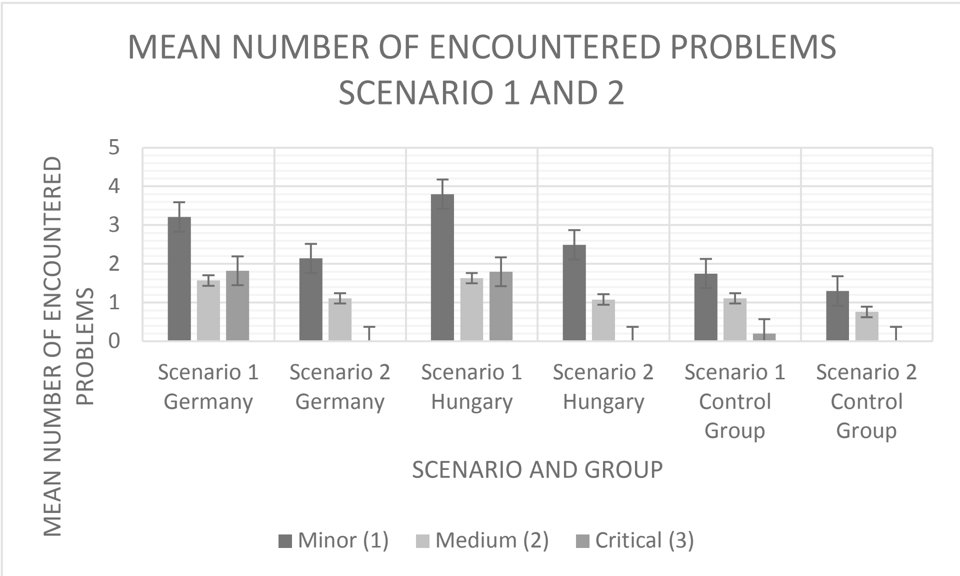


Fig. 14 – Mean number of problems encountered by the users – Both scenarios

The high number of minor problems can also be considered unsatisfactory when testing an interface. Some help from the test leader was even required in the second scenario, which followed the first and should have eliminated most problems for the users through a learning effect. This was also demanding for the elderly, with 1.98 minor and 0.98 medium problems encountered on average. Critical usability failures could not be observed in this second scenario. This can be substantiated by the effect of the learning curve (Wright T. P., 1936), (Yeh, Rubin, Hounshell, & Taylor, 2009). It can be assumed that this learning effect would eventually lead to acceptable results after three to four tests with the AusweisApp.

This is however not the solution for an interface which has to be usable the first time by users without previous experience with similar systems. A change in the process would nullify the learning effect immediately. Tests after this would produce similar results as seen above.

We can therefore state that the tested interface is not usable by elderly users without repetition or that it requires at least a deep understanding and experience with interactive e-services. Further tests based on the RSME and the ASQ also strengthened this indication. The mental effort needed for the interface was measured during testing with the standardised RSME (Rating Scale Mental Effort) created by Zijstra (Zijstra, 1993). This enabled the measurement of the subjective mental workload by comparing the user-experience to everyday tasks from easy to extremely frustrating. The results are expected. The user experience of the AusweisApp was considered by the user to require definite mental workload, scoring in the rage of 55-60 for individuals (M_{75}=58.45) on the scale of 0 to 150. This means that rather much mental effort is needed to complete the scenario. The learning effect described earlier caused the second scenario to score lower, in the region of 35-40 (M_{75}=39.72). Experience gathered through the first test lowered the mental effort needed for this scenario.

The definite method of analysing the scenario was the After-Scenario-Questionnaire (ASQ). This allowed us to compare the user experience directly through perceived attributes by the users. The three questions enabled the users to assess the scenarios and their user-experience on a scale of five scale (5 best, 1 worst)[8]. The results from the ASQ of the first scenario can be seen in table 7.

[8] This rating was used as a compromise between the German and Hungarian rating systems, and corresponds with the ratings used in the research of e-commerce websites.

Table 7 – Results from the ASQ for the first scenario based on (Lewis, 1991)

N=95	Percieved Ease of Use	Percieved Utility	Percieved Time Gain
Germany $N_{Germany}$=45	1.71	3.36	2.02
Hungary $N_{Hungary}$=30	1.97	3.27	1.77
Control Group $N_{Control}$=20	1.1	3.14	1.4

The results of the ASQ visibly show the deficits of the AusweisApp. It is considered a useful approach to G2C communication by all user groups, but user perceive it as hard to use. This was also a trend found in the answers of the control group. We consider the low scores on the first and third questions from users with substantial previous experience at using interactive Internet services to be a very strong declaration about the deficits of the application.

Results from Germany and Hungary were analogous, hence we act on the assumtion that the nationality has no impact on ASQ scores. Summing up, the scores mean that users do not perceive the system to be acceptable and would resort to offline alternatives.

The questionnaire after the second scenario gave similar results, with some improvement on the perceived ease of use. This however can be attributed to the effect of the learning curve mentioned earlier.

This also explains the higher scores in the ASQ for the second scenario in table 8. The somewhat lower results for the perceived time gain can also be attributed to this factor. Better understanding of the application enhanced the expectation of the users, but the system would not become much faster or offer more information. The scenario itself probably also lead to the higher utility: Online video-rental is a service where the users see the age-verification problem as a major nuisance and therefore any opportunity, which might give a better solution, is seen in a positive way. Such services are also used to much greater degree than online banking, in particular the usually rare event of opening of a new account.

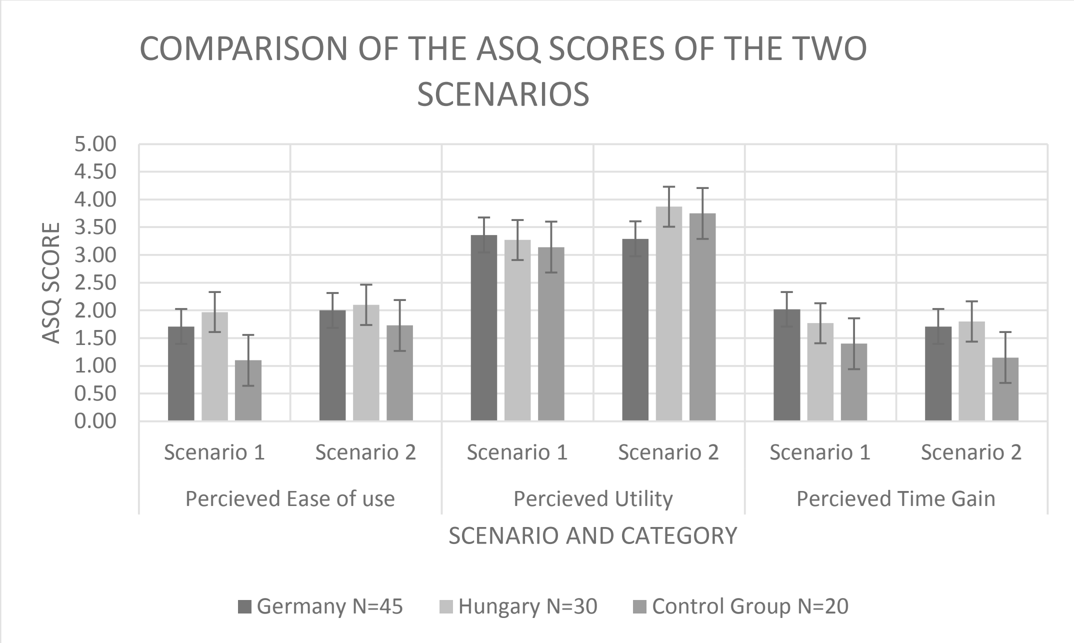


Fig. 15 – Comparison of the ASQ scores for the two scenarios, 2010 based on (Lewis, 1991)

The ASQ concluded the user tests for the second phase. The results were later used to build a framework for a better interface design, which lead to generalised solution for the optimal design of e-government systems for the ageing population.

4.2.4 Conclusion

The results not only lead to the conclusion that the interface was sub-optimal and that even the control group encountered problems, but also showed that both cohorts consider the utility of such systems favourably. It can therefore be declared that elderly users are not generally in opposition to interactive Internet applications; on the contrary, services offering good access will be accepted. On the other hand, hard to use systems will lead to fear of new technology and to a rejection of e-government as concept.

The main problem with the current interface can be attributed to the fact that it is reliant on high experience to be usable. This has been proven by the control group consisting entirely of experienced students. Users without the proper experience encountered critical problems, which not only resulted in an abort of the task, but also enhanced the required mental workload. The interface would be usable by the users after several repetitions. This means that users would learn every step by trial and error. This however is not only unrealistic, but would prove to be unreachable with current issues, which lead to avoidance of the system after the first unsuccessful trial.

We have obtained enough information about the requirements of elderly users, so we could formulate a framework which can be used to scientifically enhance the user experience. The AusweisApp will be used to demonstrate that the framework is correctly specified so as to result in systems with a higher usability.

5. IGUAN - *"Interactive Application Guidelines for the Ageing Population"*

The interviews in phase 1 and the specification of the user requirements through user testing have contributed to a better understanding of the user context of e-government applications for elderly users. The interviews have clearly shown that elderly require e-government systems and also have the curiosity to use these. This statement was fortified by the user tests, which offered an insight into the problems and the user behaviour of elderly users. These demonstrated that such systems are perceived as highly useful by this cohort, but are perceived as not usable in their current configuration.

The low usability demonstrated in the second phase with the tests on the AusweisApp lead us to the conclusion that current systems have an extremely low acceptance and cause a general rejection of interactive systems. This can also be supported by a similar study of e-commerce usage by elderly users by Birkland (Birkland, 2003). Previous research suggested several factors that affect the user's intention to use e-services. Based on the TAM (Davis, 1989), introduced earlier, these studies have demonstrated the impact of the perceived ease of use and the perceived usefulness on the user's intention to use the application. The further component of this is, besides the factors described by the TAM, the perception of trustworthiness (Carter & Belanger, 2005), which has strong impact on citizens.

There have been multiple approaches (Birkland, 2003), (Nielsen J. , 1993) to visualise these issues, thereby identifying the sources of the low acceptance of e-services by the ageing population. A number of guidelines have been developed which can be used as a starting point in the process, e.g. the design principles created by Fisk et al. (Fisk, Rogers, Charness, Czaja, & Sharit, 2009). However, there has been not much research into a persistent solution, which would increase the acceptance of e-government systems. The impact of this is not only severe, as illustrated in previous chapters, but also delays cost-reduction and might even cut certain remote locations from general citizen services. In addition, ideas for the future of e-government strongly build upon the concept of participation through e-inclusion. This is only possible by offering highly sophisticated systems to the population without excluding certain cohorts.

Multiple studies (Hipp, 2009), (Robinet, Picking, & Grout, 2008), (Bruder C. , 2008) show that elderly can benefit from certain improvements of the user interface. Acceptance can be highly increased through optimisation of the interface, thereby giving elderly a better user experience and offering systems which are not only accepted, but are preferred to current uncomfortable and time consuming offline services.

The development of the IGUAN - "Interactive application guidelines for the ageing population" – is a proposed solution for a development framework, which is engineered along the requirements of e-government systems and designed to output applications with improved utility and usability for the elderly. The idea behind the framework is built upon the structure of COBIT 4.1 (IT Governance Institute, 2007), a system audit process. Although the COBIT framework is targeted at a different aspect, the concept and structure, which unites requirements, processes and external factors, proves to be an ideal guidance to develop a framework for a similarly structured problem. This best-practice based guideline is designed to enable a higher acceptance by following the key declarations of established standards, refining and complementing these with our research conclusions, thereby creating a new, and service based user oriented guideline for e-government systems for elderly users. The IGUAN has been used in the enhancement of the AusweisApp in accordance with the newly developed best practices. This also demonstrates the suitability of the guideline by successfully enhancing the acceptance of an application with the structured approach. This serves as a proof of concept for the viability of the guideline. We designed this approach in order to follow the concept of contextual design (Beyer & Holtzblatt, 1997).

5.1 IGUAN in detail

The main consideration for the new guideline is twofold. It should not only help to develop or redesign applications in accordance with the requirements of the target group, but also provides elderly users with an easy access. In addition, it should offer aid and a visual process for developers, which leads to UI (User Interface) optimised systems. The best-practices incorporated in the guideline therefore do not have to only include the requirements for optimal usability, but also a development cycle and the general criteria, which are considered essential for system acceptability. Our user tests of the AusweisApp were used as a basis for the framework. This draft process was channelled into a more formal and standardised form to formulate a complete set of guidelines, which result in the planned improvements. This, followed by the contextual design process enabled us to implement the findings into the final application. The generalisation of the research process was possible, as it was considered from the beginning on as the first step towards a standardisation effort.

Reproducibility was also a high priority for the validity of our research, enabling an easy generalisation of the process. The framework was created with the focus on e-government and particularly electronic identification, but the guidelines can be used for any interactive Internet service which has an elderly target group.

The guideline is constructed along the three aspects shown in figure 15. These aspects are built upon the general usability improvement methods of Richter and Flückinger (Richter & Flückinger, 2010). The guideline describes how the acceptance of an application can be optimised, and how the separate aspects are interlinked with the acceptance of the Internet service. For controlling the usability improvement, IGUAN provides the three aspects, each forming a dimension of the usability aspects. The structure of the guideline enables to incorporate the internal and external aspects of the usability improvement process by integrating the core requirements of the application design with the requests of the users. This is broken down into functions which can be easily executed for an increased acceptance.

The first aspect of the guideline is the contextual inquiry (Richter & Flückinger, 2010), which provides the fundamental requirements of the target group by categorisation of the needs in three definite groups: functional requirements, design requirements and application requirements.

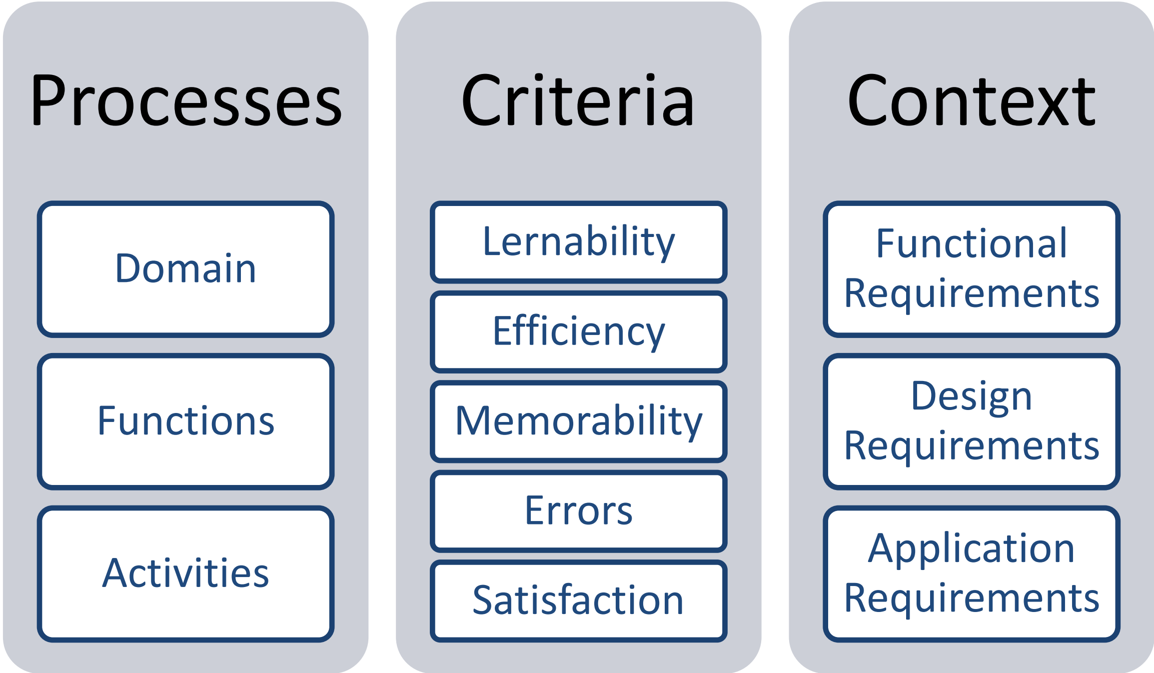


Fig. 16 – Structure of the IGUAN framework

The functional requirements include general guidance aspects and general configuration requirements. These are set by the objective of the application. These requirements cannot be modified, as they describe the internal functions and processes of the e-government service. The design requirements include generalised design issues. These are derived from the TAM (Davis, 1989) and include as focus the perceptions and attitudes towards the actual system. These have to be declared in relation to the targeted user group.

The application requirements include needs which are characteristic for the connection of the target group and the system. An aspect of e-government systems is trustworthiness, representing the citizens' perception of the credibility of the application. In addition these requirements implement factors which are characteristic for the elderly users, such as an increased font size.

The second aspect describes the criteria, which control the usability of an application. These have been selected in accordance with the experience from comparable studies in general software usability assessment (Nielsen & Loranger, 2006), (Park & Lim, 1999), (Nokelainen, 2004), and represent a link to the ISO 9126 (ISO, 2001). These criteria are the basic attribute of the interaction between the system and the user. Fisk et al. (Fisk, Rogers, Charness, Czaja, & Sharit, 2009) define five core factors, which determine the acceptability of a system. These criteria were also used in the early phases of our research into the requirements of the users. The interviews, which enabled an overview about the user requirements, were constructed along them.

The process aspect represents the main aspect of the usability improvement of the application and provides the actual best-practices for the IGUAN. These can split up into the domains, functions and activities, which directly influence the usability.

- *Domains* – The domains describe the improvement process by describing the redesign cycle with four domains. These domains are aligned with the contextual design and the usability improvement sequence cited earlier for the better understanding of system requirements of the elderly. The hybrid flow-model based on the ISO 9241-110 and shown earlier was refined to create the domain model of the IGUAN. See figure 16 for more details.
- *Functions* - The functions are the activities which result in usability improvements by specifying what is to be done to archive a better acceptance for the application. The steps, which guide the process towards the improvement, are modelled by the 23 activities.
- *Activities* - The activities are the direct actions which are required to archive a measurable usability improvement. These can be seen as the execution of the more general functions and allow for a wide customisation of the guidelines for the particular application. The activities also contain the actual methodology of the functions, thereby enabling to use the sophisticated toolset, which the actual usability improvement requires.

The functions are grouped by the domains, each domain representing a key component in the usability improvement process.

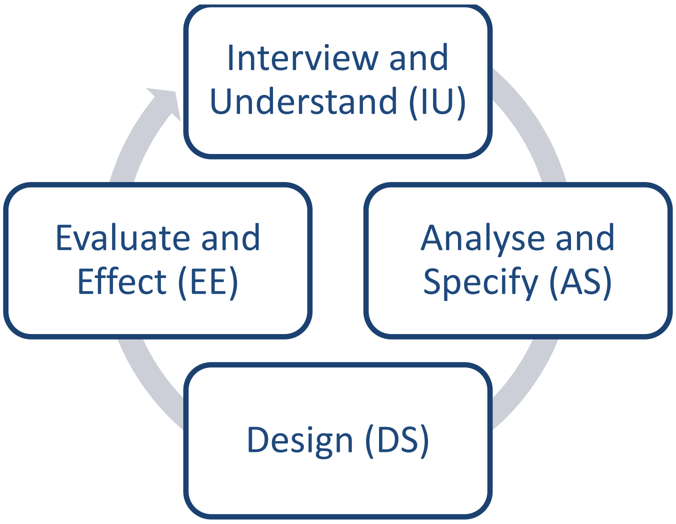


Fig. 17 – Domain model of the IGUAN

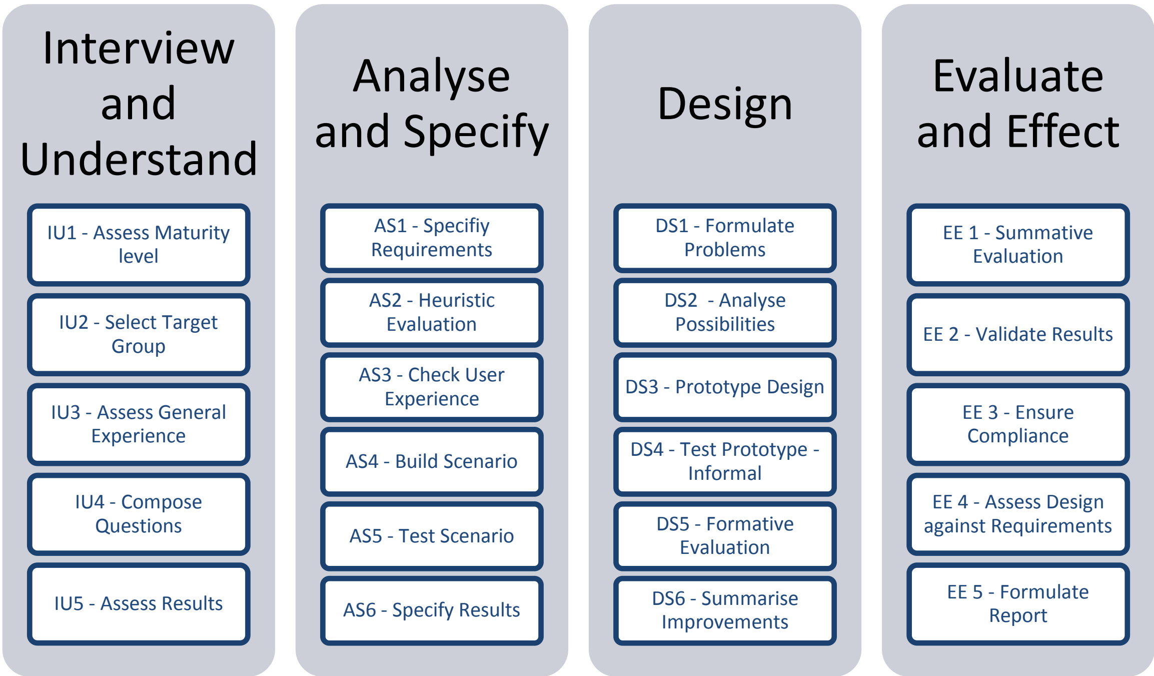


Fig. 18 – Function model of the IGUAN

Interview and Understand

This domain is needed to formulate the general attributes of the system and to understand the context. It represents the – understand the context – step of the ISO 9241-210. The objective of this domain is to identify the general aspects of the target group, and formulate the questions for the next domains by assessment of the key problems and attributes. A further objective is the planning of the general usability process. The functions of the domain are constructed with these goals in mind and help to guide the assessment by structuring the workflow.

IU1 – Assess Maturity Level

This first function is used to assess the maturity of the system in accordance with the model of the European Commission. The activities of this step include the definition and description of the overall system mission: "*What should the system do? What is the main function of the system?"*

The selected general mission has to be evaluated against the maturity model introduced earlier. This makes a categorisation possible, and enables a provisioning of the general requirements of electronic integration. The required level should be considered with the future in mind, but should not be selected higher than the actual need. Not every system needs to be fully electronic, in some cases higher electronic integration also mean less user comfort. This has to be taken into consideration when selecting the appropriate maturity level.

IU2 – Select Target Group

The selection of the target group is the second function which has to be considered. This enables the gathering of the appropriate information about the core requirements of the application. This has to be done not only on functional scale but more importantly for IGUAN on the user interface side as well: "*What are the needs of the target group? What kind of application is accepted?"*

This function gives only the approximation about the requirements. It is a heuristic method which is used to select the correct target group for the system. The set target group is inherited by further domains. Modification of this basic parameter would result in invalid results later in the process.

IU3 – Assess General Experience

This function is of high importance, as it enables the assessment of one of the core requirements the process is built upon. The assessment of the experience level of the prospective users is crucial, as it determines multiple factors of system design: *“Do the users have a computer at home? Do they use Internet applications?”*

The questions asked should be formulated along a proven scientific method, which not only offers a proven record of accomplishment but also helps to ask the correct questions. The examples shown above serve only as a basic illustration, and show merely that the general questionnaire should be based on previous experience with systems which necessitate similar user interactions.

IU4 – Compose Questions

The question formulation can be considered a further central part of the IU domain, as it controls the results to a high degree. It differentiates between success and failure in understanding the target group. General considerations should be a high comparability and standardisation of questionnaires, along with target group relevancy. In addition, questionnaires should be formulated with simplicity in mind. Typical questions should be formulated around acceptability of applications.

IU5 – Assess Results

This function concludes the interview and understand domain of the IGUAN. It represents the interviews themselves and the validation and interpretation of the collected data from the questionnaires. The activities for the interviews can include focus groups, deep interviews or other methods. The selected method has to be identical for every sub-group.

The assessment of the results enables the move towards the next domain – analyse and specify – the validation and interpretation of the data leads to the actual user tests. This function should also incorporate an activity, which includes the assessment of standards, which the target system should be compliant with. This compliance has to be assessed in the final domain, in function EE3.

Analyse and Specify

The second domain of the IGUAN represents the data-gathering on an actual system. This might be a system in use in case of a usability improvement process, or a representation of an actual system in development. In this case, the application has certain requirements, which need to be followed in order to successfully model the future system. This domain also serves as a tool to identify key problems of the system and offers answers to questions about critical issues concerning the selected group of users.

AS1 – Specify Requirements

The core of the Analyse and Specify domain is the specification of the system requirements set out by the criteria of the IGUAN. The system has to be specified in order to test and improve the system for the selected user group introduced in the Interview and Understand domain. At this point, the scope of the analysis can be developed from the results of the Interview and Understand domain. The scope of the evaluation represents the answers, which are needed for the prototyping, and should supply responses through the scenario based user tests in this domain.

AS2 – Heuristic Evaluation

As we have shown earlier, the selection of the system has another important component: the heuristic evaluation. The final system, which is to be tested, should be evaluated along certain attributes, which validate its suitability for testing with the target group: "*Is the system stable? Are the results reproducible?"*

Stability should be considered a primary concern and should be given attention. An example for the expert analysis can be seen in chapter 4.2 containing a decision-process which has proven to be well suited for the task. All technological factors should be evaluated in this function, as the selected system cannot be modified later on without compromising the results.

AS3 – Check user experience

Checking the existing experience of the test participants has been accomplished in IU3, the test group might however need an additional detailed experience analysis. The users have to be tested in order to validate the mean experience, which might influence the results in case of an abnormal distribution.

In addition, it is possible to assess the previous experience much more accurately within the scope of the application which is used for the tests. This can be done with more sophisticated methods in comparison to the general assessment done in IU3. The CLS (Sengpiel & Dittberner, 2008) has proved to be extremely useful and enabled a better understanding of the computer usage habits of the elderly.

AS4 – Build Scenario

The building of the scenario can be considered the cornerstone of the – Analyse and Specify – domain; it represents the selection of the tasks, which test the interaction of the users with the web-service chosen in AS1. Several activities should be accomplished for the optimal choice of scenarios, as the measurement of the acceptance of a system is directly proportional with the perceived usefulness of the background of the system. Activities include the research on the relevance of the tasks to the target group, as users will not regard every interaction as equally important. Good examples might be interactions which are currently used offline, and can be modelled with online processes. A further consideration is, that the scenarios should cover all important aspects of the GUI, which is to be tested. It should further contain all relevant functions and features.

AS5 – Test Scenario

The testing of the scenarios with the target group is the main source for the results in this domain. The testing has to be done in accordance with the methodology, which is also selected in this function. This methodology is also to be used in the design domain, in DS4 and DS5. This ensures that the results are valid and also comparable. Activities for this function contain the actual testing with a larger group of participants. The number of tested users might vary to a certain degree in conjunction with the system parameters, but as a rule of thumb, the minimum sample size should be oriented in accordance to other studies (Turner, Nielsen, & Lewis, 2004). In addition, control groups should be considered in certain scenarios. These can be used to verify the connection between the problems encountered by the target group and other groups outside the target. This can yield valuable information about the impact of the GUI problems in relation to the usability criteria set out earlier in this thesis.

AS6 – Specify Results

This function can be considered as a connector to DS1 describing and formulating the usability problems found in AS5. If the system and the scenarios were selected successfully, the outcome should be illustrated by a comparably high-perceived usefulness of the application, but a lower acceptance. Key problems should also be identified through AS5, as it is imperative for a successful usability improvement process later in the design domain.

Design

The design domain represents the direct usability improvement in the process. This is the actual delivery of requirements identified in the Analyse and Specify domain, although it includes a control through the iterative prototyping cycle, which pinpoints improvements for key issues. The output of the domain should be an almost user ready application prototype, which has a higher acceptance and an improved GUI tailored for the requirements of the target group.

DS1 – Formulate Usability Problems

This function is strongly interconnected with the AS6, and represents the continuation of the specification of the results by quantifying the actual problems. This includes the registration of each problem in accordance to Nielsen and Loranger (Nielsen & Loranger, 2006). This makes it possible to grade usability problems in accordance with their seriousness and prioritises their ramification for the users. This can further be used in DS2 to specify the key problems of the interface.

DS2 – Analyse Improvements

The identification of the key issue of the interface should be followed by the evaluation of possible solutions for these problems. Resolutions should be based on theories or practical usability improvement approaches, which have been verified by comparable studies. New and unproven theories can be formulated, but should be tested in the iterative process to validate the significance of the improvement. This function also represents the planning of the implementation priority based on the insight from previous verified ideas and the knowledge acquired in the previous function, DS1.

DS3 – Prototype Design

The prototype design is the first step of the development cycle, which is the key component of the prototyping domain; it is the basis for further research and improvement, creating a route inside the usability improvement process. The development cycle should contain the elements for a successful GUI redesign illustrated earlier. Prototype design and verification should be used at the beginning and end for each iteration with informal and formative tests included in between. This procedure ensures the feedback from each prototype, and makes the iterative development possible through constant improvement.

DS4 – Test Prototype – Informal

The informal testing is a method which enables usability testing of minor components of the global interface on a smaller scale, which has not only the advantage in costs, but also makes a direct comparison of concurrent approaches possible without the requirement of a higher number of participants (Mark et al., 2009). Informal tests only need a small number, three to eight users, who are presented by small snippets of the actual interface in different functional or design layouts. Small-scale scenarios or tasks enable the users to voice their feedback on the possibilities (Krug, 2010). The comparison of different on-screen-keyboard designs illustrates this advantage clearly. It enables a cost effective method to assess the requirements for this GUI element without the need for large-scale full usability tests.

DS5 – Formative Evaluation

This function represents the main flight of the user tests of this domain. It is comparable to the AS6, and should be conducted with the comparable scenarios and an identical methodology in order to guarantee validity and comparability. As with the DS4, the evaluation should be done for each prototype which is part of the design cycle, as this will enable an insight on usability improvements. This will deliver results which can be used in the refinement of the GUI. The validity of the implemented enhancements of the interface should be verified by tests with a larger group and assessment of the effects on the control group.

An important activity is the verification of the comparability. This is to be archived through the maintenance of an identical research methodology and duplication of the scenarios from AS5. Compliance to this is essential for retaining the validity and significance of the results.

DS6 – Summarise Improvements

The summarisation of the results is the basis for the summative evaluation in the Evaluate/Effect domain. This is needed for a successful comparison of the initial system with the application resulting from the AS and PT domains. Activities in the DS6 function should include a categorisation and summation of the uncovered issues and their solutions: "*Which are the key improvements? How can they be categorised?"*

It should be noted that the summarisation of the results will also be needed for an assessment against the requirements in EE4, therefore the categories formulated have to be in line with the general problems of the target group uncovered in the Interview and Understand domain.

Evaluate/Effect

The evaluation of the results and the assessment of the effects of the new system are the final elements of the usability improvement process. This domain represents the summarisation of the results by assessing them against the requirements and results acquired in the other three domains. This element of the process also incorporates the verification of the results against standards which should be adhered to, and where the compliance was incorporated in IU5. The correct execution of this domain should therefore result a system which is more accepted and has a measurably higher usability than the original application, which was input into the IGUAN. It also results in a comprehensive report, which points out the possible solution. By following the guideline, the results should be reproducible, valid and compliant with the requirements of the target group.

EE1 – Summative Evaluation

The final evaluation of the results and the prototype from the DS5 concludes the evaluation of the application. A summative method is needed for a global insight and a valid comparison with the original systems, which was input into the IGUAN. This evaluation can be executed with a complete array of users, including control groups consisting of previously tested individuals. This can offer direct data about the improvements, as these participants have experience with the task, but not with the new system. Assessment of the results in comparison to new participants without any previous contact to the system might further validate key aspects modified and offer a broad insight into effects of usability improvements for the target group. The summative evaluation should be made with the same methodology as any previous tests and the formative evaluations in the improvement process. The ASQ method is a tool, which can be used in both the smaller scale tests and the large-scale summative evaluation, thereby guaranteeing valid and comparable results, which can be used to formulate a report.

EE2 – Validate Results

The validation of the results is not only imperative to guarantee the consistency of the data and the output system, but also to verify the representativeness of the data acquired from the user tests. Even large sample sizes and the effort to collect the data from a seemingly representative group of users might result in unrepresentative data, as over-representation of highly experienced users might falsify the results. Validation through modelling should therefore be considered as a verification of the results. We will demonstrate this by using a modelling tool, which provides quantitative predictions of the tasks, thereby verifying the representativeness of the sample.

EE3 – Ensure Compliance

This function is built into the domain to evaluate the output system against the standardisation requirements set out in IU5. Activities include comparison of requirements in relation to the actual system, thereby ensuring the compliance.

EE4 – Assess Design against Requirements

The assessment against the requirements is one of the key aspects of this domain. It is imperative as it validates the direction of the improvements. The main activity of this function is the interpretation of the data from the summative evaluation – EE1 – against the requirements of context. This enables a comprehensive overview of the new system and the improvements and validates the expected higher acceptance of the new application. This function is also the basis for the reporting in this domain.

EE5 – Formulate Report

The final function of the IGUAN guideline is the reporting, which includes the output of a consolidated report to visualise the improvements of the application by highlighting the enhancements in acceptance and usability. This is required in order to summarise the general improvement and can be used as a platform for further research on the system.

IGUAN therefore provides best practices and guidelines and does not offer rigid design principles for e-government systems. These would make universal application for the vast number of different systems with vastly different objectives and goals impossible. This function and activity based approach combined with the criteria and the context make a workflow possible which can be used for a wide variety of applications and offer guidelines to make e-government more usable for the ageing population.

6. Produce Design Solutions – Proof of Concept of the IGUAN

This phase of the research serves as a sample, which shows that the IGUAN is applicable for its designed purpose. In order to confirm the applicability of the framework for a generalised approach, it is used on an e-government application. The successful execution of the usability improvement process through the use of the guideline serves as a proof of concept. The software for this step was the AusweisApp which was previously used as sample application in the second phase of our research. The reasons for the selection of the AusweisApp are two-fold. The reuse of the already known software environment, processes and user tasks enables to remove distortion of the data stemming from the use of a new application. In addition this enables to handle the processes and the tasks of the software environment as a constant, which does not interfere with the results. This also makes it possible to focus on the effect of the structured approach of the usability improvement process.

The reuse of the AusweisApp processes and tasks also enables a comparison between the original and the improved client. The direct comparison also makes the proof of concept possible as it shows a measurable usability difference between the original application and an application redesigned by using the framework.

As we have shown in the previous chapter, the IGUAN is built on three aspects, which incorporate the internal and external requirements of the system in the criteria and the context pillar. In addition, the processes pillar enables the implementation of the usability improvement process.

The improvement process has to be started by filling out the Interview and Understand domain of the process aspect. This enables the identification of the requirements which are necessary for the enquiry of the other two aspects by identifying the criteria and the context of the respective system. The criteria are divided into five main aspects which control the usability of the application and cover the general user requirements for the particular system.

- *Learnability* – The application should enable the users to learn how to use it. Users should become measurably more proficient by using the system multiple times.
- *Efficiency* – The application should allow users to achieve the intended task within reasonable amount of time.
- *Memorability* – The application should be designed in a way that users remember how to use it. This implies that the effort of relearning following non-use should be minimal. This criterion is important, as the typical usage patterns for e-government systems is rare usage.
- *Errors* – The application should be constructed in a way that user-errors while using the system should be recoverable. Feedback should be given to the user.
- *Satisfaction* – The application should give the user a pleasant user experience.

The second aspect which has to be taken into consideration concerns the context of the system. This contextual inquiry provides the functional requirements of the target group by categorisation into functional, design and application requirements.

- *Functional requirements* – The main functional requirement of the application is the electronic identification of the users.
- *Design requirements* – The design requirements for the application include aspects, which makes it pleasurable to use. It should be in line with the design language of the operating system it is running on. It should appear sophisticated, business-like and modern. The design should be usable by the elderly.
- *Application requirements* – The application should comply with the following aspects: Elderly users should perceive the system as trustworthy and credible, as the system handles personal data. Font size, feedback and logic of the interface should be designed according to the needs of elderly users.

After the criteria and the context have been identified, the main part of the guideline can be used. As we have shown previously, the domains of the process pillar contain the functions required for the usability improvement process. These functions have to be executed in order to fully comply with the guideline and output a system with an improved usability. The Interview and Understand (IU) domain is the first which has to be used in the usability improvement process. It contains the aspects of comprehending the general processes, the requirements of the users and compose the questions for the next domains by assessing the key problems and attributes. At this stage of IGUAN, there is no actual system selected, only a general category.

The IU1 to IU5 have been executed through the first phase of the research. This has given information about the general goal, the maturity, the target group and their general experience. The interviews of the first phase, which assessed the acceptance of e-government by the elderly in Germany and Hungary, comply with IU4 and IU5 functions. These gave a broad understanding about the requirements of the prospective users and their expectation about the system category.

The next step in the system inquiry is the Analyse and Specify (AS) domain. It represents the data-gathering on an actual system, thereby enabling a deeper comprehension of the specific interaction of the target group with the application. It also offers tools to identify key problems of the system by a step-by-step system requirements analysis. The starting step of the domain is the AS1 function. This step is essentially the selection process of a suitable system for the previously selected target group, which will be further used in the analysis. We have accomplished this step by selecting the AusweisApp in second phase of the research. The selection process should be supported by a reproducible method with criteria based on the identified requirements. This was done with a decision process shown in chapter 4. This system was then evaluated in AS2 along previously selected benchmarks and resulted the selection of the AusweisApp.

As a following step (AS3), the user experience has to be analysed against the selected system. This can also be seen in chapter 4. For this system we used the CLS (Sengpiel & Dittberner, 2008), which has proved to be extremely useful and enabled a better understanding of the computer usage habits of the elderly.

The composition of the scenarios and the testing (AS4 and AS5) are the next two phases in this domain. These enable the direct testing of the system against the requirements identified earlier. These can also be found in chapter 4. The AS6 function describes the found problems and enables the prototyping domain.

The third domain includes the functions, which enable the improvement of the AusweisApp. In order to successfully execute the prototyping, the problems which were broadly uncovered by the AS5 have to be formulated into measurable tasks. Subsequently, these tasks have to be prototypable and comparable so that a scientific measurement is possible.

In our case the following problem areas have been uncovered by the DS1 and DS2 functions, which should be solved in development cycle of the design domain:

- As seen in the requirements, elderly have special needs for the legibility of text in a computer interface. The GUI of the AusweisApp was found not to comply with these requirements leading to several otherwise avoidable errors. This aspect has to be investigated in the design domain. Papers by Nebe, Zimmermann and Paelke (Nebe, Zimmermann, & Paelke, 2008) and the ISO 9241-110 also consider this as a key quality indicator for the interface.
- A critical semantic problem, which was identified by the AS5, was the structure and display of the service provider. The current design, which can be seen in chapter 4, is not only unsuitable, but also conflicts with the criteria of learnability. It is not comprehensible to the user and implicates that users do not read the provided information. This problem has to be resolved to improve the overall usability of the client.
- A further requirement, which can be derived from the criteria, is the need for a progress indicator. This stems from the criteria memorability and efficiency. The system has to offer a feedback for the user about the made progress and the pending steps. This has to be implemented in a way, which is compliant with the identified requirements and further criteria. A study by Sharit et al. also suggests, that the use of a graphical aid that provides a visual model of the progress improved the ability of older people to navigate complex systems (Sharit, Czaja, Nair, & Lee, 2003).

- An issue, which was observed in AS5, was the unclear separation of the data and the PIN input into the client. The criteria memorability and learnability were harmed by the implementation of both steps on the same screen. Wright, Blythe and McCarthy (Wright, Blythe, & McCarthy, 2006) found that implementation of different logical steps of a process into a single GUI screen will have a negative effect on the usability of an interface.
- An aspect, which was derived from a functional requirement, was the need for a screen-keyboard for the input of the PIN.
- Characteristics, which also have to be taken into account in the usability analysis are the application and design requirements. The system has to provide a satisfactory user experience and should be perceived as credible. The AS5 of the AusweisApp shows that the criterion satisfaction is not met. Fisk et al. (Fisk, Rogers, Charness, Czaja, & Sharit, 2009) suggests that this can be attributed to the lack of feedback in the system. This problem is a crucial aspect, which is implemented into the design phase.

  The design requirements also have to be fulfilled by creating an application, which is in line with the design language of current operating systems.

These aspects have to be included into the DS3, DS4 and DS5 functions of the Prototyping domain. These have to be executed along an iterative cycle for each problem category described. This iterative method was chosen as this approach represented the operation of the ISO 9241-210 for the design solution. Without this intermediate systematic process, it could not be guaranteed, that the new GUI is indeed superior to the initial AusweisApp.

The other two dimensions of the usability assessment, the procedures and tasks, were left unaffected. This decision described in the previous chapter was made in order not to compromise the comparability of the results. This application development cycle built upon the IGUAN enables the incorporation of a route inside the usability improvement process containing the crucial functions of the prototyping domain.

The Prototype Design (DS3) function was used based on the issues uncovered in the requirement specification phase. This step produced prototypes, which included the improvements. The input for GUI modifications was twofold; either from verified research results and guidelines, or from problems uncover in the specification phase.

The Informal tests (DS4) enabled a pre-validation of the improvements and could be used to test key components without a complete formative evaluation of a new prototype. This was used multiple times in the iterative approach to filter out concurrent implementations of GUI elements. The test groups for these informal sessions were kept as small as possible, but were kept significant by using five test participants (Nielsen J. , 1993). A larger number of participants would not have made the results significantly better. In accordance with Nielsen and Landauer (Nielsen J. , 1993), a test with five users will uncover 85% of the usability problems of a system. Some authors (Krug, 2010) concluded that five participants are not enough for significant results, in case of the informal tests however, the expected result is insight, not validation. In this case, the five participants are enough for determining the best implementation of GUI elements (Barnum, et al., 2003).

The Formative evaluation (DS5) of the completed prototypes with a larger fifteen user record enabled a deeper insight into the problems of the prototypes. The procedures used in this step represented the same quality as in the second phase of the project. The smaller sample size was a compromise between data significance and time constraints; a sample of fifteen can however be considered adequate for usable results in these intermediate tests. This can also be based on the findings of Nilsen and Landauer. In addition the increase of the number of participants to fifteen results in an improvement in data confidence (Faulkner, 2003).

The proof of concept of the data acquired by the third step was used to build a new and improved prototype through validation. Each modification of the new GUI was validated through expert analysis and comparison to standards. These gave the foundation for the next prototype, which was developed in the application development step of the next iteration.

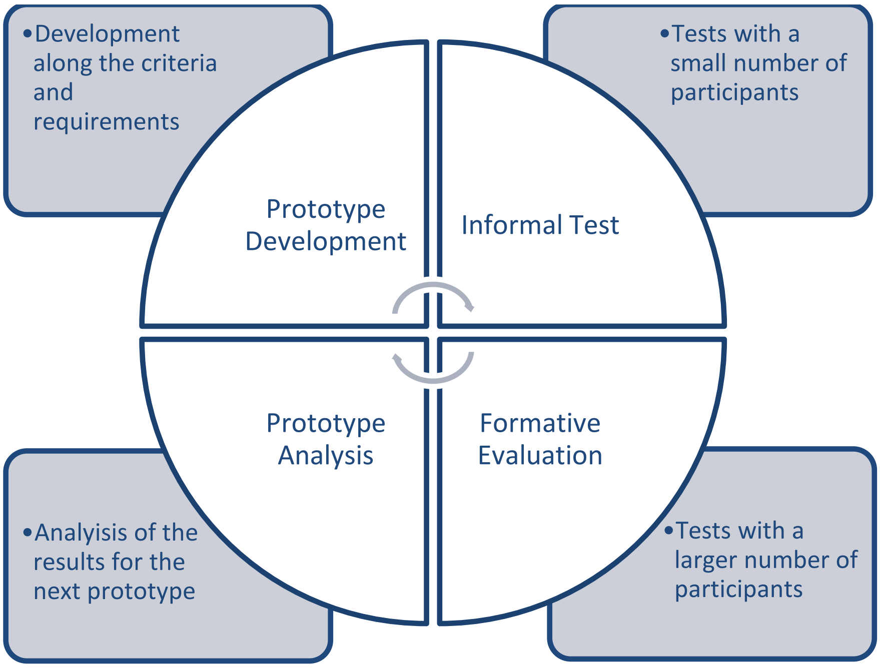


Fig. 19 – Iterative prototyping and testing (Source: based on (Boehm, 1986))

6.1 Iterative Prototyping and Testing

The prototyping process for the AusweisApp was developed along the criteria stated above; ideas and developments were created not through heuristics, but through data derived from the Analyse and Specify domain through crosslinking these with the requirements and criteria defined by the analysis. These were either proven or confuted by the tests later in the process. The measurement of the usability of the prototypes was based on the ASQ_{pe} score of the ASQ. As we have observed in the previous chapters, the ASQ is a good indicator of the perceived usability of a system. As the initial system already possessed a comparably high ASQ_{pu} score, the significance of this for the usability of the system was deemed negligible and only measured in the summative evaluation. The ASQ_{pt} score is only descriptive for a larger scale test and was therefore used as secondary indicator.

6.1.1 Prototype 1 – Assessment of the legibility

The very first issue, which had to be taken in account was the problem with the legibility of the interface. This is one of the major requirements for the software declared through the context aspect of the IGUAN, and it is one of the key components which can be considered as key quality indicator of the interface (Nebe, Zimmermann, & Paelke, 2008). As stated in the ISO 9241-110 from 2006 and even the older ISO 9241 Part 3 (1992), easy readability is one of the crucial aspects of good usability. Developing a user interface especially for population groups with a higher percentage of presbyopia, as the elderly, means that the GUI has to be adjusted to this requirement (Lohr, 2006). Contrary to expectations, Chadwick-Dias et al. (Chadwick-Dias, McNulty, & Tullis, 2003) found that text size did not significantly affect the usability of websites – the outcome was attributed to increased amount of space the larger text needed and the requirement to scroll. Nevertheless, elder-specific recommendations set font-size to match 12-14 points on paper. Bernard et al. (Bernard, Liao, & Mills, 2001) report that the reading speed and user preference for sans-sherif fonts is 12, whereas sherif type is read faster at 14 points, on the whole, sans-sherif font was preferred by the elderly.

This statement was tested with addendum that the higher usability has to be accomplished in accordance with the core statement that the system has to be still usable by the general population. The first prototype was therefore used to find this optimal font size, which was tested with younger $(N_y = 5,\ M_y = 25.8\,years,\ SD_y = 4.08)$ and older participants $(N_o = 5,\ M_o = 72.6\,years,\ SD_o = 5.74)$ with an informal evaluation. The original AusweisApp features a font size (Arial) of 10 points, this was seen as problematic by 53 of 75 test participants in the second phase of the result.

The tests were conducted with the first screen of the prototype featuring the information about the certificate. This screen contained multiple lines of significant text which should be read by the users for a secure G2C communication with the AusweisApp. The initial 10-point font was increased in steps of 1 point for both test groups up to 16-points with the acceptance recorded. The client was displayed on a 22” LCD with a Full-HD resolution (1920x1080 pixels).

Results pointed out that the acceptability will decrease for younger users if the general font size is increased beyond 14 points. The 10, 12 and 14 point fonts were tested in a function reduced prototype against each other. The results from the t-test can be seen in table 8. The consensus was that a good compromise between legibility for the elderly and an appealing interface is a lettering with a font size between 12 and 13. The interface was regarded somewhat more pleasing with a font size of 12, therefore this was used in the later prototypes, and emphasized text was displayed with bold characters, as on the original AusweisApp. Semantic changes were not made in this prototype. These findings are also consistent with the results of Chadwick et al. and Bernard et al. (Chadwick-Dias, McNulty, & Tullis, 2003), (Bernard, Liao, & Mills, 2001).

Table 8 – ASQ Scores in relation to the font size in the prototype

Prototype	**Control Group (N_y=5)**		**Test Group (N_o=5)**	
	ASQ_{pe} Score	SD	ASQ_{pe} Score	SD
10 point font	1.13	0.74	1.23	0.41
12 point font	1.3	0.98	1.61	0.68
14 point font	1.64	1.5	1.89	1.9

6.1.2 Prototype 2 – The certificate problem

A semantic problem in the second phase was the issue with the certificate which displays the validity of the provider in the G2C communication. This is also consistent with the criteria of learnability. The main problem with the implementation of this in the original AusweisApp is that the users do not read the exorbitantly long text containing the essential information. As shown earlier, the information is presented in unfamiliar terms for the elderly. This is not comprehensible to the user without deep knowledge about the processes of the system (Rosson & Carroll, 2002). The reaction to this is an instinctive click on a button taking the user to the next screen without having read the text. This problem has been demonstrated with a paper prototype in which the information displayed has been modified and displayed "*We are a criminal gang and will steal your personal information*" instead of the actual details of the Fraunhofer FOKUS. Out of the 30 candidates tested, none has read this security related information in detail, all have clicked instinctively on the button.

The main problem, which is caused by this issue is a serious security flaw, which can cause the users to send their sensitive personal information to parties not authorized. The paper prototype was tested with 5 younger ($N_y = 5,\ M_y = 26.8\ years,\ SD_y = 3.42$) and 25 older participants ($N_o = 25,\ M_o = 72\ years,\ SD_o = 4.47$). When asked about this, they admitted not reading such long legal information, and the younger participants compared it to EULAs (End User Licence Agreement), which they usually do not read.

This problem was solved by replacing the element by a green OK field, which visualised the certificate in a comprehensive and easier to read package. This concept was based on the research of Rosson & Carroll on visual metaphors (Rosson & Carroll, 2002). It was tested in accordance with the development cycle shown earlier in this chapter. The results have shown that this solution can be used in future prototypes, as it improved the comprehensibility of the task enormously. The new method was tested against the AusweisApp style with two test groups $\left(N_{AusweisApp} = 12,\ M_{AusweisApp} = 72.6\ years,\ SD = 5.74\right)$, $(N_{P2} = 13,\ M_{P2} = 70.6\ years,\ SD = 5.77)$ and a control group of students for both prototypes. The perceived ease-of-use in the ASQ of the first scenario improved from 1.59 points to 1.68 points and the mean number of critical problems decreased to 1.5. The results from the tests with both test groups (younger control group and target group) can be seen in table 9.

Table 9 – ASQ Scores in relation to the certificate display

Prototype	Control Group			Test Group		
	N	ASQ_{pe} Score (Mean)	SD	N	ASQ_{pe} Score (Mean)	SD
“AusweisApp” Style	5	1.8	0.118	12	1.50	0.522
Prototype 2 (Fig. 19 for more Information)	5	2	0.15	13	1.92	0.494

Fig. 20 – The OK-Field, which complemented the certificate information field in the first screen of the prototype, Source: screenshot, taken December 2010. "AusweisApp" Prototype 2

The hypothesis H_1 that the improved certification information field has a significant positive effect on the ASQ scores was tried with a significance test for the test group, where px defines the hypothesis for the prototype x.

(1) $$\boldsymbol{H_1}: \boldsymbol{\mu_{P2}} \geq \boldsymbol{\mu_{AusweisApp}}$$

The t-test (see Appendix C) indicates that H_1 can be accepted at significance level of 5%. The empirical results from the t-test is $t_{emp2} = 2.083$, which is larger than $t_{crit}(0.95,23) = 1.714$, therefore the improved certification field has a significant positive effect on the ASQ scores.

As an additional measure, the impact of the certificate display on the total mean number of critical problems was also tested. This was done with 5 participants. The hypothesis H_1, that the number of errors is influenced by the new layout, was tried ($N_{AusweisApp} = 5, M_{AusweisApp} = 1.805, SD_{AusweisApp} = 0.435$), ($N_{P21} = 5, M_{critical21} = 1.5, SD_{P21} = 0.35$).

(2) $$\boldsymbol{H_1}: \boldsymbol{\mu_{P2}} \neq \boldsymbol{\mu_{AusweisApp}}$$

The t-test (see Appendix C) indicates that H_1 can be accepted at significance level of 5%. The empirical results from the t-test is $t_{emp\_crit2} = 2.556$ and $p = 0.034$, therefore the improved certification field has a significant effect on the total number of critical errors.

6.1.3 Prototype 3 – The Progress Indicator

An additional logical requirement, which was observed in the second phase of the research, was the need for a progress indicator, where the process steps are visualised. This problem also affected the efficiency of the GUI, thereby damaging one of the key criteria of usability. A study by Sharit et al. (Sharit, Czaja, Nair, & Lee, 2003) found that the use of a graphical aid that provides a visual model of the progress improved the ability of older people to navigate complex systems. The AusweisApp featured this indicator in form of a field on the left side of the GUI. This method of progress visualisation is not the customary layout and was misinterpreted by the elderly users as a menu pane.

Designing GUI elements where the function is not clear leads to serious problems with the usability of the interface (Tidwell, 2005). In this case, the function of this interface element misleads the users and has caused several problems in the second phase of the research. Users misidentified the element for a clickable navigation field, and did not comprehend its function as a progress indicator. To rectify this, the prototype 3 explored the optimal method to implement a progress indicator in applications for the elderly.

Several GUI elements were created which were tested separately from the complete prototype. These were designed along the ISO 9241 and in accordance with previous research in this field by Nielsen (Nielsen J. , 1993). The early designs featured modified indicators on left side. This was evaluated in order to retain the overall characteristic of the AusweisApp. This had to be omitted from later prototypes, as results from the tests $(N_{P31} = 14\ \ M_{P31} = 71.25\, years,\ SD = 5.03)$ showed that the confusion of the users was caused not only from the GUI element itself, but also from its location on the left side of the main information. Rosson and Carroll also encountered this problem and described that interpretation of GUI elements is strongly affected by their location (Rosson & Carroll, 2002).

Table 10 – ASQ Scores in relation to progress indicator

Prototype	Test Group		
	ASQ_{pe} Score	SD	N
"AusweisApp"	1.6	0.805	75
Prototype 3.1 Progress indicator on the left	1.79	0.426	14
Prototype 3.2 Linear progress indicator	2.15	0.555	13

Our further research and prototyping enabled a new approach to the progress indicator and resulted a GUI element, which proved to be superior to the original one. This element was created along the premises from the studies introduced above (Nielsen J. , 1993), (Tidwell, 2005). As seen in the comparison, the prototype 3.2 built with this indicator $(N_{P32} = 13,\ M_{P32} = 70.22\ years,\ SD = 5.40)$ resulted in a significantly higher ASQ_{pe} score.

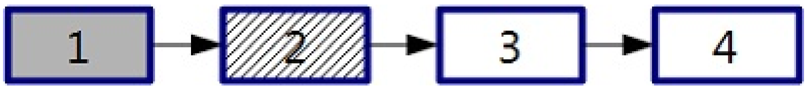


Fig. 21 – Final prototype 3 progress indicator design, Source: screenshot, taken December 2010. AusweisApp Prototype 3

The hypothesis, that the linear progress indicator results significantly better ASQ_{pe} scores was tested with a t-test:

(3) $$\boldsymbol{H_1: \mu_{P32} \geq \mu_{P31}}$$

The t-test (see Appendix C) indicates that the H_1 can be accepted at a significance level of 5%. The empirical result from the t-test is $t_{emp2} = 1.943$, which is larger than $t_{crit}(0.95,25) = 1.708$.

We can therefore declare that the linear progress indicator results significantly better ASQ_{pe} scores than the progress indicator on the left. The mean number of encountered critical problems per test candidate decreased from the prototype 2 by a further 0.5 points, measuring a mean of 1 critical problem per test candidate. Refined derivatives of this indicator were used in every following prototype.

6.1.4 Prototype 4 – Logical separation of the steps

An issue, which was not only observed in the second phase, but was also seen as a concern in the heuristic analysis, was the unclear logical separation of the second and third input step of the application. The logical separation of user input steps in the interface is a decisive requirement and is also a key criterion.

In case of the AusweisApp, both input steps (personal data and PIN) were displayed to the user on the same screen, which turned grey, but was still readable. This implied to the users that the input elements were both active and could be interacted with.

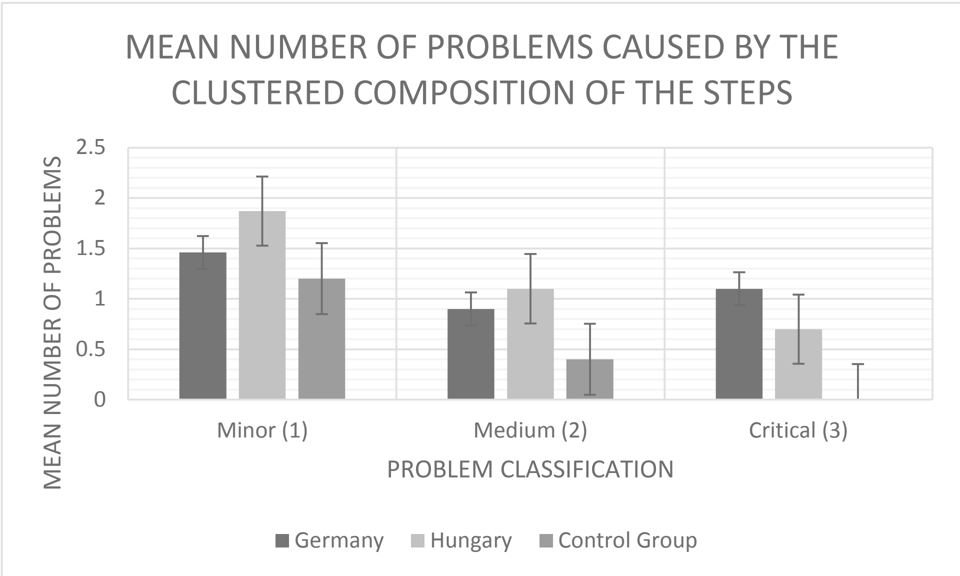


Fig. 22 – Mean number of problems caused by the clustered composition of the AusweisApp for both scenarios

It was observed that users were confused by the multiple input options, which were only active sequentially. This critical problem discovered in the tests implicated that the logical separation of system steps is essential to fulfil the criterion of learnability in an application. This is not only a key aspect of the IGUAN guideline, but it is also supported by findings of other studies in HCI (Wright, Blythe, & McCarthy, 2006).

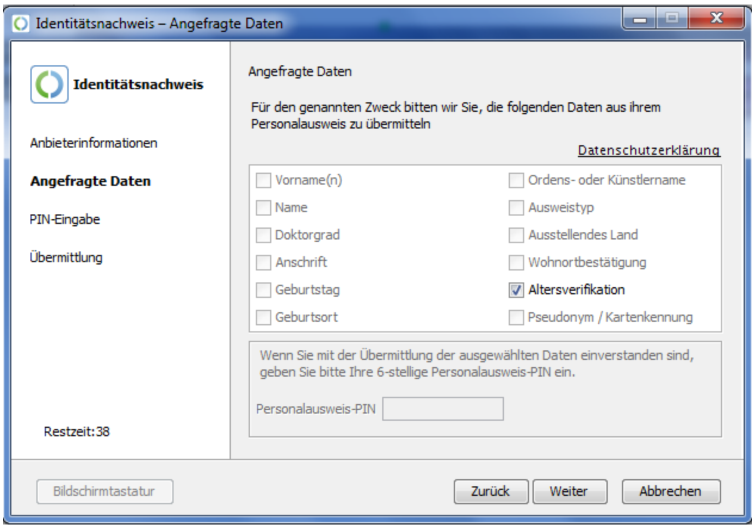


Fig. 23 – "AusweisApp" Version 4.1 – Data and PIN Input, (Source: Application created by the Bundesministerium der Innern/Fraunhofer FOKUS, 2010)

We deducted a hypothesis from these results, which declared that the solution of this problem will have a significant positive impact on the acceptance of the application. This served as the basis for the prototype 4, where the two steps were put on different GUI screens. Our hypothesis $H_1$was tested with this new prototype:

(4) $$\boldsymbol{H_1: \mu_4 \geq \mu_3}$$

The prototype 3 (μ_3) represents the integrated layout, where the steps are clustered on a single screen. The prototype 4 (μ_4) denotes the new approach, which should score significantly better at ASQ_{pe} according to the findings of Wright et al. (Wright, Blythe, & McCarthy, 2006).

The statement that the prototype 4 enabled higher ASQ_{pe} scores was tested at 5% significance. A t-test was used, as it gives accurate results for smaller samples. The component-based evaluation of the new prototypes with the separated steps was based on a test with five candidates$(N_{P32} = 9,\ M_{P32} = 70.22\ years,\ SD = 5.40), (N_{P4} = 5,\ M_{P4} = 72.6\ years,\ SD = 3.72)$ with the same premises as the tests with the previous prototype featuring an identical layout, except for the integrated steps.

Table 11 – Results from the t-test

Prototype	N	ASQ_{pe} Score for ease-of-use	
		Mean	SD
Separated	5	2.2	0.837
Integrated	9	2.11	0.782

The t-test implicates that H_1 cannot be accepted at $\alpha = 5\%$, as the significance is $p = 0.846$. (see Appendix C for more details) This means, that the new layout had no significant effect on the ASQ_{pe} scores, and the higher scores can be explained as being coincidence. We suspect that the inconsistency of the results of the t-test with the high number of encountered problems is a result of the small sample. This problem might require further research to fully understand the connection between the logical composition of a process and the usability of the GUI. We decided however, that the somewhat higher ASQ_{pe} score for the separated design meant that elderly users will regard the separated design to be equal or better.

An additional problem, which was observed in this step, was the issue of the checkboxes. The second phase of the research showed that multiple participants found these to be challenging to hit and select. Due to their relatively small size, where the mouse has to be aligned to successfully select the option, they are problematic to use by a target group whose motor and visual skills might be deteriorated. The hypothesis formulated from this observation was that larger selection tools should enable a higher acceptance.

This issue was also encountered by other research projects. Sayago et al. concluded that radio-buttons and check boxes are usable by the elderly (Sayago, Sloan, & Blat, 2011). This was however disproved by the tests of the AusweisApp in late 2010. The check boxes were deemed to be too small by about 45% of the test group in Germany and Hungary ($N_{S2} = 75,\ M_{S2} = 69.68\ years,\ SD = 6.38$). It can be hypothesised that while previous experiments were conducted with full-screen web pages, the AusweisApp is a window-based client, which requires the users to focus on a small part of the screen.

The prototype 4.2 represents a proposed solution for this, where larger and more contrasting selection boxes are used. These turn to colour inverted when selected. This idea was based on the Microsoft Windows User Experience Interaction Guidelines by Microsoft (Microsoft, 2010).

This was tested with a t-test, where the previous prototype (μ_4) represented the check-box variant of the interface. The new (μ_{41}) prototype featured the selection boxes ($N_{P4} = 5$, $M_{P4} = 72.6\,years$, $SD = 3.72$), ($N_{P41} = 5$, $M_{P41} = 71.8\,years$, $SD = 3.86$).

Table 12 – Results from the t-test

Prototype	N	ASQ Score for ease-of-use	
		Mean	SD
Selection boxes	5	2.4	0.548
Check boxes	5	2.2	0.837

The t-test implicates that the selection boxes have no significant effect on the ASQ_{pe} score. The mean number of encountered critical problems further decreased however compared to the prototype 3, and measured 0.4 ($N_{P41} = 5$, $M_{P41} = 71.8\,years$, $SD = 3.86$). The decrease in the encountered problems suggests that the logical separation has an impact on usability, and that selection boxes are a useful solution for window-based applications.

6.1.5 Prototype 5 – The Screen-keyboard

The German Federal Ministry of Interior formulated the need for a screen-keyboard as a requirement, which increases security of the AusweisApp. This can be translated into an application requirement, which the application has to be compliant with. It is also an imperative component of secure PIN based applications.

These two factors make it a principal candidate to be included in the evaluation. This component was first included in the series 5 prototypes, with several on-screen-keyboard layouts tested. For these tests, the last series 4 prototype was modified with each layout and tested with eight participants from the focus group. The first design tested was a sequential layout based on number arrangement of standard computer keyboards. This arrangement can also be found on some online banking websites, i.e. www.deutsche-bank.de. The layout is 1 2 3 4 ...etc. The number zero can be on either end, it is not standardised. A delete button is on the right end of the sequence. The second layout tested was the calculator layout, commonly found on calculators and computer keyboards with a number block. This layout features the number 1 in the left bottom corner and incorporates a larger zero button. Additionally, this layout may also include a “<-” button for deletion of a number. The third layout was the telephone configuration, found on fixed and mobile phones, PIN input fields on ATMs (Automatic Teller Machine), and on some alarm systems. Such keypads feature the number 1 in the top left corner, the number 0 is a normal button and the layout features two special buttons in the bottom line. These were labelled “X” for complete delete of the input and “<-” for delete of one number. The tests were conducted in sequential order with an identical task. The comparability was warranted by the ASQ_{pe} and by the time measurement of the input of the 6 digit PIN of the participants ($N_{P5} = 8,\ M_{P5} = 71\ years,\ SD = 3.60$).

Table 13 – Results of the on-screen keyboard evaluation

N=8	Mean ASQ ease-of-use	Mean Time Requirement
Sequential	1.4	194 s
“Calculator”	2.2	143 s
“Telephone”	2.6	130 s

The results supported strongly the telephone configuration. This is used on devices, which need a PIN input thereby creating a familiar environment for the users, where the input is independent from the hardware device. The results in table 13 lead to the conclusion that this PIN field input layout should be used for optimal acceptance of an on-screen-keyboard.

6.1.6 Prototype 6 – Form and Feedback

IGUAN also contains the design requirements of the application. These have to be assessed for a complete proof of concept of the guideline. These components not only incorporate the visual aspects of the design, but also some visual feedback, which the user requires for a satisfactory user experience. The requirement for an integrated help feature was not only observed in phase 2, but is also consistent with the criteria of learnability and satisfaction declared in the IGUAN. The need for this element was not only hinted by the comments of the target group, but is also supported by multiple papers - "*built-in help functions increase the usability of devices significantly*" (Baumann, 1999). According to Fisk et al. (Fisk, Rogers, Charness, Czaja, & Sharit, 2009), older adults are also more likely to make errors when interacting with systems lacking a help system. This study also shows that error recovery is more likely in such systems. The prototype 6 was therefore used to verify this in the case of e-government systems. An integrated help function was implemented directly into the application, which displayed the relevant information on the respective step. This feedback feature gives not only general information, but in case of the data selection it also provides the user with a short description of the data, which is about to be sent. This was necessary as not all data categories are self-descriptive. The second phase of the research clearly indicated that the target group did not always understand terms such as proof of age or pseudonym.

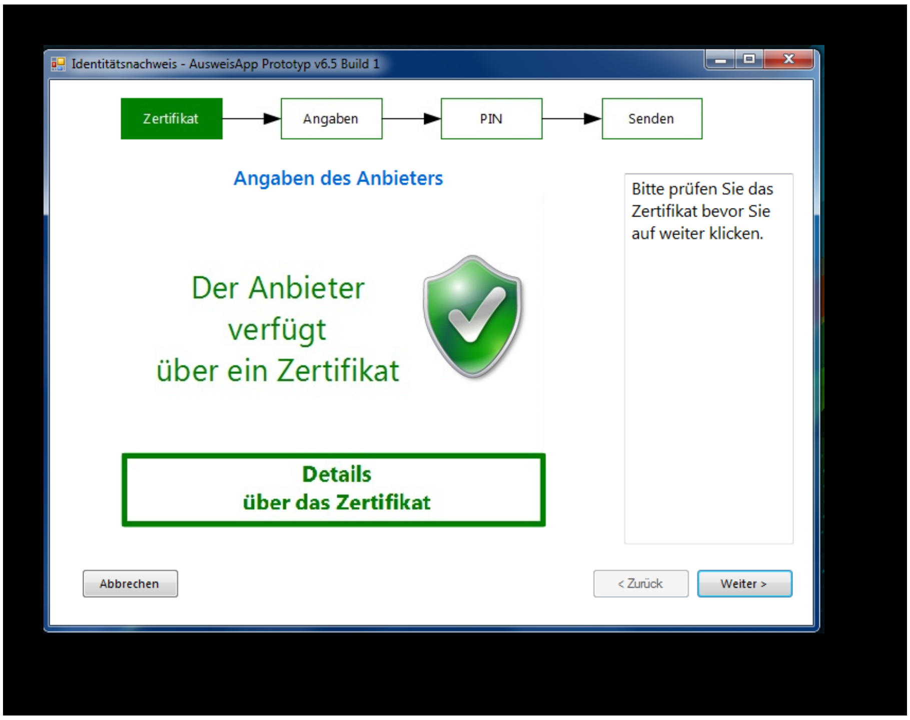
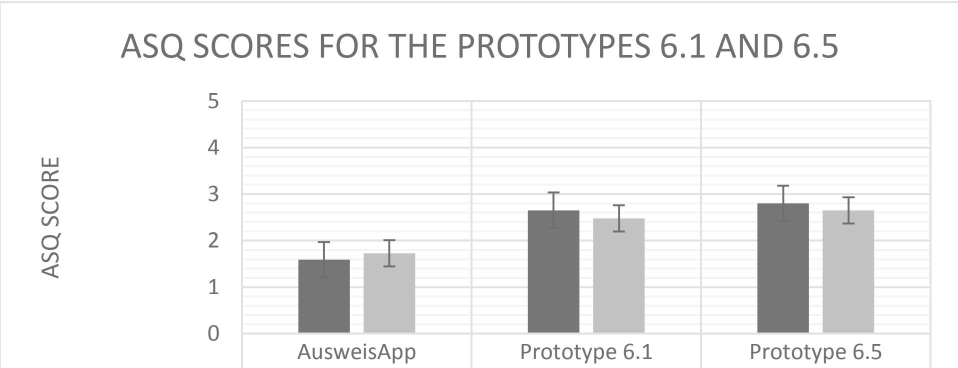


Fig. 24 – Integrated feedback feature in the prototype series 6.5, screenshot taken in July 2011

The new prototype was tested in accordance with the methodology described, with the assumption, that the feedback should increase the respective ASQ_{pe} score and decrease the time needed to complete the tasks. The test group was composed of eight participants. $(N_{P61} = 8,\ M_{P61} = 71\ years,\ SD = 3.60)$

The results $(ASQ_{pe61} = 2.65, ASQ_{pt61} = 2.48, T_{61} = 126s)$ showed a strong increase in the ASQ_{pe} score, this increased from 2.4 points in the prototype 4 by 0.25 points. The ASQ_{pt} increased by 12% from 2.2 to 2.48 points. The mean time needed to complete the first task was 126 seconds, which was anticipated, as the new element increased the amount of information, which had to be read. In a further step, the prototype was redesigned in accordance with the "Windows User Experience Interaction Guidelines" (Microsoft, 2010), which helped to create an aesthetically more pleasing application, more in line with the design language of current operating systems. The theory that the aesthetics of an application will increase the usability was based on statements of Hassenzahl and Sutcliffe (Sutcliffe, 2010), (Hassenzahl, 2010), who experimented with psychological factors of user engagement. - "*Users favour applications for serious use with a focus on content* ..." (Sutcliffe, 2010) - but for most users "*For initial impression, aesthetics may be important...*" (Sutcliffe, 2010). This factor was taken into account in the prototype 6.5, which was tested with a smaller control group of seven test participants $(N_{P65} = 7,\ M_{P65} = 70.2\ years,\ SD = 2.92)$ to verify that aesthetics are a component of the design requirements of the IGUAN and have an impact on the acceptability of an e-government application. The functionality of this prototype remained identical.



	AusweisApp	Prototype 6.1	Prototype 6.5
ASQ Ease-of-Use	1.59	2.65	2.8
ASQ Time-Gain	1.73	2.48	2.65

Fig. 25 – ASQ scores for the prototypes 6.1 and 6.5 in comparison with the AusweisApp

This theory has been verified, as ASQ scores demonstrate a higher perceived ease-of-use for this prototype ($ASQ_{pe65} = 2.8, ASQ_{pt65} = 2.65, T_{65} = 120s$). The ASQ_{pe} increased further to 2.8 points from 2.67 points in prototype 6.1. The mean time needed to complete the first task decreased to 120s.

6.1.7 Prototype 7 – GUI Final

Prototype 6.5 can be regarded to be a streamlined layout, where issues uncovered in the previous research phase have been resolved. The GUI final prototype was therefore developed as a final step. This represents the Summarise Improvements function of the IGUAN seen in DS6. The prototype series 7 includes GUI improvements, which were used to fine-tune the final application prototype for the summative evaluation in the Evaluate/Effect domain. These improvements correspond to the three requirement categories set out by IGUAN. Functional requirements incorporated according to the input from previous studies with similar applications. Design improvements were built upon the facts acquired from diverse guidelines described earlier. Finally, application requirements for the elderly have also been included in accordance with the needs uncovered earlier in our research.

User interaction supported some rearrangement of the buttons; the “Abort” button was repositioned into the lower left corner of applications. Additional modifications were the implementation of a visualisation of the step number and the relocation of the data selection fields in the second step. This enabled a feedback field nearer to the actual information, thereby providing a faster and more efficient GUI. The final design for this can be seen below.

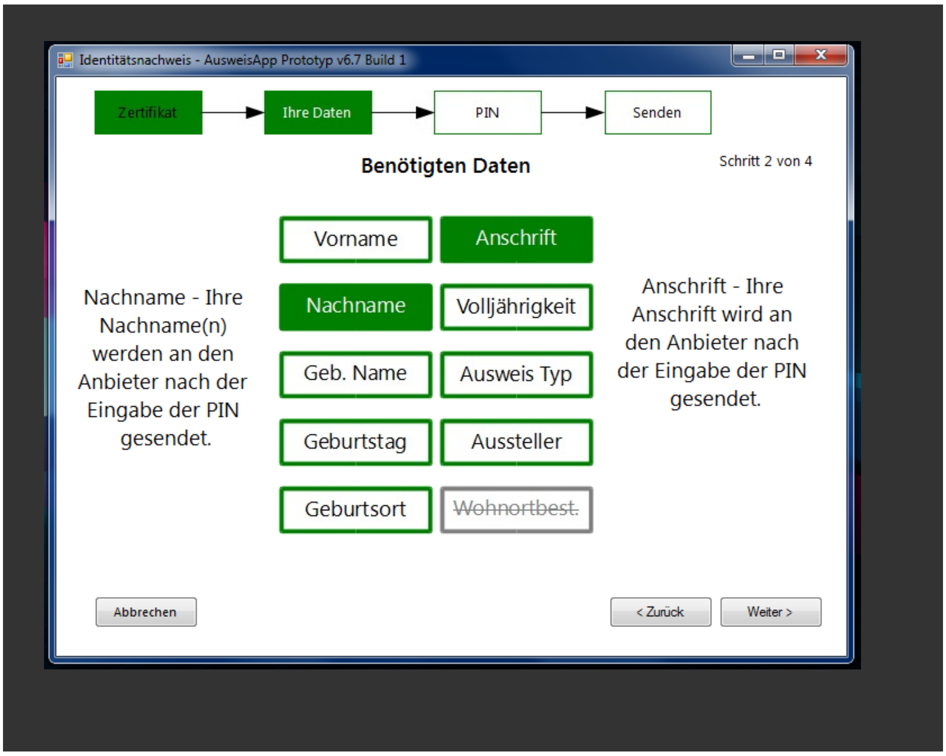


Fig. 26 – Optimised and final GUI prototype for the “AusweisApp”, screenshot taken in September 2011

The interface now features less colours, the use of only white and green signals trust and consistency to the user, and the black 12 point font allows an easy to read interface. The progress indicator developed in the prototype 3 was also further enhanced featuring a description of the actual steps in place of the numbers in previous versions. This enabled a better orientation of the users. This prototype was tested before the final summative evaluation. This step was used to be able to measure the improvements in comparison to the prototype series 6.5. $(N_{P65} = 7,\ M_{P65} = 70.2\ years,\ SD = 2.92)$ $(N_{P7} = 7,\ M_{P7} = 70.8\ years,\ SD = 3.76)$

Table 14 – Results from the t-test

Prototype	N	ASQ Score for ease-of-use	
		Mean	Deviation
Prototype 6.5	7	2.8	0.837
Prototype 7	7	3.0	0.707

We tested the hypothesis H_0which states that the final GUI (prototype 7) leads to identical or lower ASQ scores than the prototype 6.5:

(5) $$\boldsymbol{H_1: \sigma_{65} \neq \sigma_7}$$

An F-test was used to determine, whether the prototype 7 results higher ASQ_{pe} scores, than the prototype 6.5.

(6) $$F = 5.74, F(6.6) = 4.28$$

H_1 can therefore be accepted at significance level 5%, prototype 7 results a lower variance. We can therefore assume that prototype 7 results significantly higher ASQ_{pe} scores than prototype 6.5.

6.2 Summary of the improvements

The DS6 is the final step in the design domain of the IGUAN. This concludes the evaluation of the interface prototypes and summarises the improvements, which were made through the improvement cycle of the DS3, DS4 and DS5 functions. In this step the AusweisApp is tested against the prototype which was created by adhering to the IGUAN. This enables a first comparison between the input and the output of the IGUAN. This is then followed in the next domain by a summative evaluation, which is also more conclusive.

The summary of the improvements in this function is only usable as a small scale conclusion as the sample size of the new design is not comparable to the sample size of the AusweisApp.

The theory, that the improvements have an impact on the perceived usability of the interface and thereby on the acceptance, can however be formulated and tested with a t-test.

Table 15 – Participants of the t-test (AusweisApp/Prototype 7)

Interface	N	Age (Mean) years	Age (SD)
AusweisApp	75	69.68	6.38
Prototype 7	7	70.80	3.76

Table 16 – Results from the t-test

Interface	N	ASQ Score for ease-of-use	
		Mean	Deviation
AusweisApp	75	1.6	0.805
Prototype 7	7	3.0	0.577

A significant difference in acceptance can be implicated by the results: $\overline{X_{P7}} > \overline{X_{A41}}$ at a significance level of 5%. The results show that prototype 7 has a significantly higher ASQ_{pe} than the original AusweisApp. We can therefore declare that prototype 7 has a significantly higher acceptance. The empirical result from the t-test is $t_{emp} = 4.481$, this is larger than the critical t-value $t_{crit}(0.95, 80) = 1.664$.

The comparability of prototype 7 with the AusweisApp is ensured by implementing only modifications which were derived from the requirements and criteria identified in the respective functions of the IGUAN. We based the core requirements for elderly users on conclusive results of comparable research projects. These were shown in detail in the previous chapter. We confirmed these requirements through our user tests with elderly users.

These factors enabled to uncover several core design deficits, some of which are generally applicable to all users. It seems that users of both groups considered the AusweisApp as not trustworthy. This was caused by the unappealing user interface. This also caused minor and medium problems for both groups. In contrast, several critical problems we identified are unique to the computer interaction of the elderly users. One of the main issues for elderly users was the need for special accessibility, which was translated into GUI modifications such as the larger font size or the adaptation of selection boxes. The use of high-contrasting colours and inversion for selections was also implemented to enable an easy access for elderly users. The more pleasing GUI design which was created in accordance with the studies of Hassenzahl (Hassenzahl, Experience Design, 2010) and

Sutcliffe (Sutcliffe, 2010), also contributed to a significantly better acceptance of the application.

We have also analysed the impact of the logical structure of the application on the acceptance. The improvements, which we made on the structure of the application, were based on the logical refinement of the visualisation of the process. Our conclusion was that there is a need for a logical separation of definite process steps. Users with low experience poorly understand steps of the process where multiple logical steps are integrated into one visual display. This observation was also supported by relevant studies with comparable e-services and the results of the according prototypes. The results support the theory that separate logical implementation of functional steps will increase the acceptance of an application. Studies by Dickinson et al. (Dickinson, Eisma, & Gregor, 2003) and Lumsden (Lumsden, 2007) also concluded that logical separation of interaction steps is one of the key components of improving usability.

An additional aspect, which is supported by the results, is the need for a feedback function in an e-government application. The lack of feedback strongly affected the acceptance and the efficiency of the GUI. ASQ_{pe} scores increased significantly after adding a feedback function. It enabled better comprehensibility of the process and decreased the perceived complexity of the application, thereby increasing the acceptance by elderly users. It also gave the user information about expressions which they were unfamiliar with. In our case, data fields with unclear terminology were explained. A study by Shneiderman (Shneiderman, 2003) that tests multi-layered interfaces supports these results.

The more intuitive information input resulting from the implementation of the selection boxes and the on-screen-keyboard for the PIN input resulted a significant increase in acceptance. The intuitiveness of GUI elements was verified with multiple concurrent designs, which were evaluated simultaneously. The concurrent GUI variations were based on findings of similar projects and on observations made on appliances with similar components. ATM PIN keyboards were used as a model for the on-screen-keyboard of the application. We discovered that such input is preferred by elderly users mainly because of the similarity to the PIN input process in already experienced processes.

The implementation of standards enabled us to build prototypes which are compliant to current principles. The prototypes thereby contain established and known solutions for common problems offering elderly users a familiar environment.

The summarisation of the improvements derived from the prototype 7 concludes the domain and leads to the Evaluate and Effect (EE) domain of the IGUAN. This final step describes the evaluation of the results obtained through the previous domain and assesses the effects of these improvements on the usability by comparing the results with the input application on a larger scale.

6.3 The Summative Evaluation

As required by the IGUAN, a summative evaluation (EE1) is employed to validate prototype 7 against the initial situation. This summative user test concluded the data gathering and enabled a provision of the results. To ensure a valid assessment of the results against the data gathered in the second phase of the research, the methodology remained unchanged, the sample was gathered with similar premises and the tasks were left identical. This enabled a direct comparison between the original, unchanged application and the application, which has been put through the guideline. The summative evaluation was conceived to produce valid and replicable results, which validate the effect of the improvements possible by using the IGUAN guideline.

6.3.1 The Task

The tests were based on the scenarios shown earlier; the first scenario tested was the opening of a bank account with the electronic interface; the second task was the age verification at the online video rental. The electronic interface was the prototype 7, respectively in German and Hungarian. Internal difference to the initial test with the AusweisApp was the absence of a connection to the servers of the Fraunhofer FOKUS and the communication of the client with the card and the connected reader. This was simulated by placing the card on the connected reader and launching the client with a Java Script on a HTML website. The participants were asked to click on a button embedded into the website, thereby launching the application. This ensured that there was no noticeable difference in the process from the user standpoint, making the simulation comparable with the AusweisApp. The questionnaires before and after the test were identical to the ones previously used to guarantee a comparability of the results.

6.3.2 The Sample

Table 17– Test sample of the summative evaluation

	N	M	SD
Hungary	35	69.82	5.62
Germany	40	69.7	3.97

The number of test participants was based on the premise of comparability with the previous tests. Candidates recruited from Hungary ($N_{S1H} = 35,\ M_{S1H} = 69.82\ years$, $SD_{S1H} = 5.62$) and from Germany ($N_{S1D} = 40.\ M_{S1D} = 69.7\ years,\ SD_{S1D} = 3.97$) respectively. The gender distribution was $F = 34, M = 36$ with an equal distribution of genders in both sample groups. The place of residence was comparable to the previous tests with the majority of the participants from the federal states of Berlin and Brandenburg in Germany, and from the metropolitan region of Budapest in Hungary. To ensure the comparability of the results, test participants were new to the system; no test subject had any previous experience with the test setup.

6.3.3 The Variables

As the methodology of the tests was identical to the previous assessments, there were no significant changes in the selection of dependent variables. The test environment (AusweisApp and prototype 7) should be considered the independent variable. The aim of summative evaluation was to show that the improvements made by the IGUAN resulted in an application, which has measurably higher usability than the client currently offered. The dependent variables are as in the previous steps, the ASQ scores, the mean number of encountered errors and the time required to complete the tasks.

6.3.4 The Procedure

The procedure was identical to the previous tests as alterations would impair the comparability of the results. The test procedure consisted of the questionnaire and the user tests of the prototype.

6.3.5 The results

The GUI improvements were well received by the elderly. Earlier user evaluations implicated that the idea of the AusweisApp results a high utility score. This statement can be deducted from the high perceived utility score of the ASQ (ASQ_{pu}). The ease-of-use (ASQ_{pe}) and time gain (ASQ_{pt}) was perceived as low for the intial AusweisApp as the respective attributes scored comparably low. This however has changed in this test set-up. The elderly users perceived this new GUI of the prototype 7 superior in every aspect.

To ensure that the favourable results do not stem from a higher experience of the sample, we measured the previous experience of the participants in accordance with earlier tests. The CLS was used to measure the previous experience of the test group with computer systems.

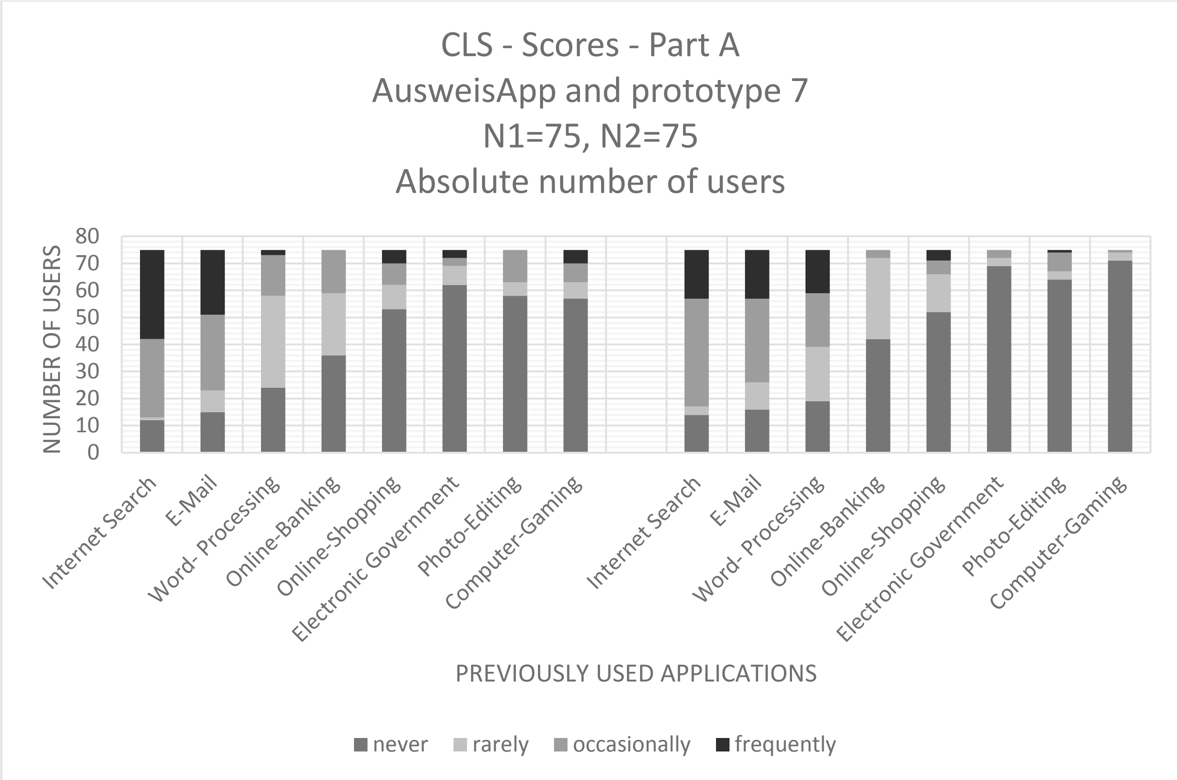


Fig. 27 – Results from the Part A of the CLS for the initial "AusweisApp" and the Prototype 7, Source: CLS (Sengpiel & Dittberner, 2008)

The results of the CLS were comparable with that of the sample of the original AusweisApp. As shown in figure 27 the two samples have comparable experience; the impact of the previous knowledge should therefore be equivalent for both.

The improvement of the GUI and the higher acceptance of the system were illustrated by the higher respective ASQ scores and the lower RSME rating in comparison to the original AusweisApp. The RSME rating was N_{75}=41.98 for the online banking and N_{75}=31.10 for the film rental application. This not only shows that the system is generally better accepted, but also that less mental workload was needed for the improved application, which not only enables a faster use, but also reduces the frustration with the interface.

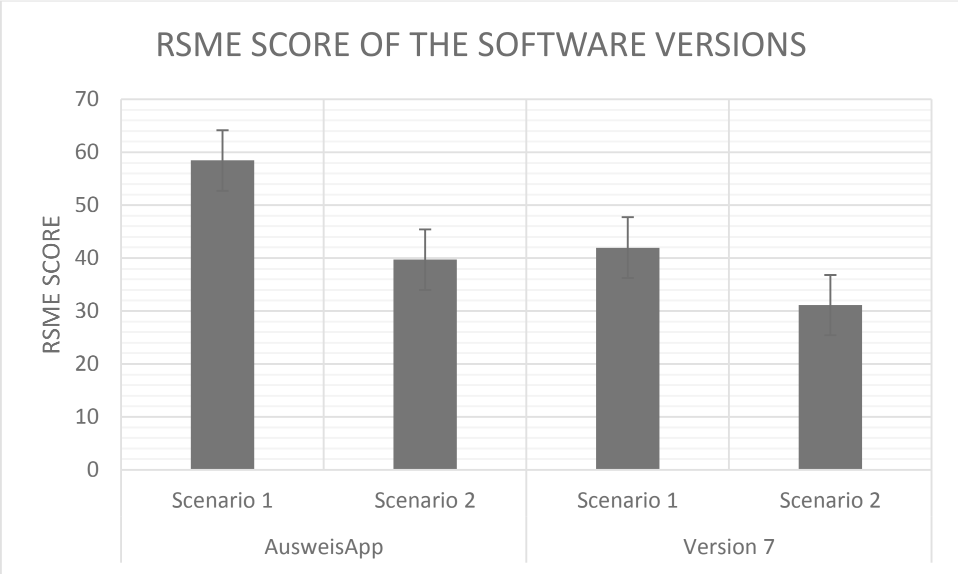


Fig. 28 – Results from RSME analysis of the AusweisApp and the Prototype 7 for both scenarios. Source: (Zijstra, 1993)

The lower mental effort needed to use the application contributes to a higher willingness to use the e-government system instead of offline services.

The mean number of encountered problems could be reduced significantly, with no critical problems observed. This enabled every participant to complete the process without the likelihood of abort and supports the hypothesis that better interfaces contribute to a higher number of successes and to a better user experience.

Table 18– Average number of problems encountered per user – Scenario 1

N = 75		Minor (1)	Medium (2)	Critical (3)
Germany N=40	AusweisApp	3.21	1.57	1.82
	Prototype 7	1.42	0.50	0.00
Hungary N=35	AusweisApp	3.80	1.63	1.80
	Prototype 7	1.35	0.70	0.00
Mean	Prototype 7	1.39	0.60	0.00

The mean ASQ scores were within the same scope as for the small-scale formative evaluation with eight test participants. The $(ASQ_{pe} = 3.0)$ and the $(ASQ_{pt} = 2.8)$ can be considered good for an application where the general task and the process was unknown to the participants before the tests. The low previous experience with such systems was also compensated by the easy-to-use interface design. The $(ASQ_{pu} = 3.40)$ can also be considered a good score. This indicator improved only slightly as the original AusweisApp was already seen as useful by the test participants in phase 2 and scored adequately for the ASQ_{pu}. The improvement of 67% and 38% respectively in the ASQ_{pe} and ASQ_{pt} scores signifies therefore that the new GUI is considered to be easy to use; and the perceived time gain is more dominant when using the new interface. These improvements in the perceived ease-of-use combined with the already good perceived usefulness of the e-government service will ensure a better attitude towards usage and enable a higher actual system use (Davis, 1989).

Table 19– ASQ Scores prototype 7, summative evaluation

Score	Scenario 1 – Online Banking	Scenario 2 – Online Video Rental
ASQ – Perceived Ease of Use	3.00	3.12
ASQ – Perceived Time Gain	2.80	3.20
ASQ – Perceived Utility	3.40	3.74

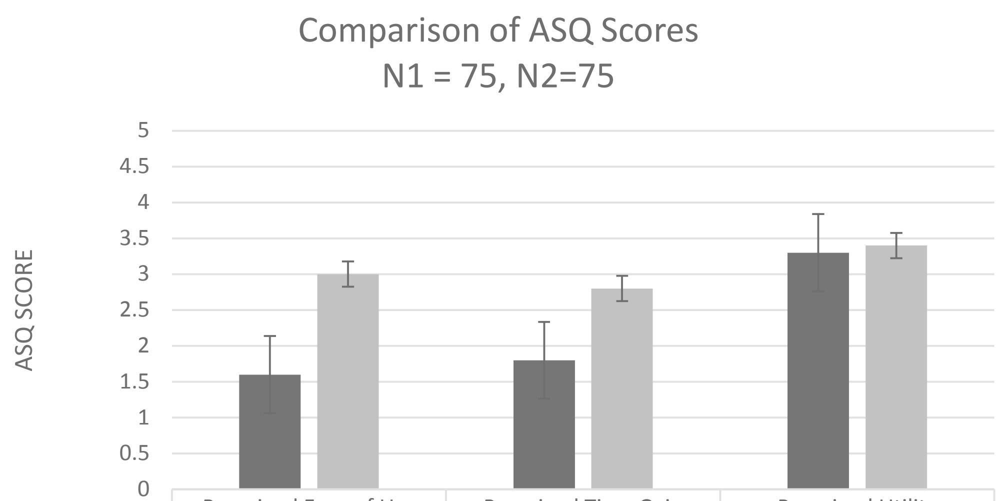
The higher acceptance of the IGUAN enhanced GUI can also be shown through a t-test of the results. This shows that the new interface, which was created along the uncovered requirements has a significantly positive impact on the ASQ_{pe} score of the GUI. The hypothesis that the $\mu_{prototype\ 7} \geq \mu_{original}$ can be accepted through the following analysis of the ASQ_{pe} scores: $(N_{p7} = 75,\ M_{p7} = 69.76\ years,\ SD_{p7} = 4.6)$, $\left(N_{original} = 75, M_{original} = 68.27\ years, SD_{original} = 6.35\right)$

The ASQ_{pe} score means differ significantly between the AusweisApp and the prototype 7, as the empirical result is $t_{emp7} = 10.130$, which is larger than $t_{crit}(0.95{,}148) = 1.645$.(for details, see Appendix C)

This means that $\mu_{prototype\ 7} \geq \mu_{original}$ and the H_1 hypothesis can be accepted.

The conclusion from this multi-level analysis is that the improved interface does increase the ASQ_{pe} score significantly. This also indicates that the independent variable (the interface) has significant direct influence on the acceptance of e-government systems and that it is possible to enhance systems in a way, so that elderly perceive them as ease to use.

The direct comparison of the ASQ scores between prototype 7 and the AusweisApp in figure 29 also shows the improvements, which result from the GUI improvements.



	Perceived Ease of Use	Perceived Time Gain	Perceived Utility
AusweisApp	1.6	1.8	3.3
Prototype 7	3	2.8	3.4

Fig. 29 – Comparison of ASQ Scores – Prototype 7 vs. AusweisApp

6.4 Comparability of the results

Previous chapters have shown that the methodology used in the research enabled the development of a prototype system which is considered more usable and is also supporting a higher acceptance by the elderly users in comparison to the initial application. This conclusion was reached through an expert review based on a heuristic evaluation and user tests with the targeted user group. The direct testing was conducted twice with 75 participants from Germany and Hungary recruited from regions within the countries with a comparably high population density, but also with a higher than average education.

This might have resulted higher than average computer and Internet experience, which might have influenced the results. This validation of the results (EE2) is a function implemented into the IGUAN to filter out possible problems which might stem from an unrepresentative sample. In case of the AusweisApp, modification of the sample in relevance to the educational statistics of the respective countries might generate a sample, which represents the average elderly user in the designated country. This can be translated to candidate recruitment from statistically less privileged regions thereby balancing the higher experience of some users and cancelling out any effect which might endanger the representativeness of the results.

This is however impossible to accomplish with elderly users. As observed by most researchers, it is almost impossible to acquire test participants in the age group 65+ with low educational background or from low-income households. The need for technology is not extraordinarily important or a high priority for users without enough income to purchase a computer or Internet connection. It can therefore be declared that although the medium to highly educated, medium to high-income users are over-represented in our research, this will not invalidate the results per se, as these elderly users have the prerequisites to use e-government and other interactive Internet applications. Low income or lowly educated elderly people will not be able to use e-services without communal access to these. The solution of this problem would need however more research and is also not a psychological or computer science question, but a sociology issue.

The only problem for our research, which might be caused by the over-representation of the medium to highly educated is their statically higher Internet aptitude and experience. This might lead to overly optimistic estimates. This issue is problematic as it would make the prototype and thereby the generalised best practices unfeasible and ultimately unsuccessful.

We have removed the possibility of this issue by creating a model representing the interaction of a user with an optimal experience with the GUIs. We formulated the hypothesis that our users had sub-optimal experience and represented the possibly most representative user group available.

To visualise that the sample had less skills than the optimal user, software based modelling of the optimal user was chosen. The CogTool software was used, which allowed the modelling of the user interaction and evaluated the predicted human performance with a cognitive model (John B. E., 2009). This software enabled to create a storyboard of the designs, incorporating the demonstrated task and creating a cognitive model for the optimal user. The software supplies a prediction of the time required for the user with an optimal experience to complete the scenarios. This prediction made it possible not only to validate prototype 7 by validating our hypothesis that our users were not optimally experienced, but also made a direct comparison between the different prototypes possible. It helped to visualise the improvements between the different prototypes as the model gives a directly comparable estimate on the time needed to complete the task with the model. The CogTool produced these quantitative predictions of human performance through three steps:

- *Interface prototyping* - The prototypes shown earlier in chapter 4.3 were directly imported into CogTool with every step representing a possible interaction. The states of the GUI were implemented into a storyboard. The storyboard of the interface prototypes thereby represents a direct translation of the original user interactions. The frames themselves contain widgets, which represent the interactive elements of the GUI that users can manipulate. Keyboard and mouse interaction is also integrated into the model.

- *Task modelling* - The two mock-up tasks described in chapter 4.2 were implemented into the CogTool model and were used for the task modelling for each interface prototype. This enabled the direct relationship between the results from the user tests and the estimates from the CogTool. The application uses the storyboard to produce a computational cognitive model of a highly experienced user (John B. M., 2009). CogTool is based on the Keystroke-Level Model (Card, Moran, & Newell, 1980) and also make use of the ACT-R (Adaptive Control of Thought – Rational) model (Anderson & Lebiere, 1998) which helps to understand the human cognition.
- *Projection* – CogTool creates this with the storyboard and the task calculated with the Keystroke-Level Model and ACT-R. CogTool also creates a visual comparison and script of the prediction, which can be used in the analysis of the results.

A storyboard with the corresponding frames was created for every prototype, and the projection model was computed with the two tasks introduced and used in the user tests. The attributes and the general design of the tasks enabled us to use the same storyboard of prototypes for both tasks without major modification necessary. This has also contributed to the comparability and general validity of the cognitive modelling.

The frames for the model were built from screenshots of the prototypes with every interactive element, consisting of buttons, input fields or selection boxes, being represented by a widget. These widgets enabled the computation of the optimal model by selecting the path of interactions of the task. Generally, the script of the model was kept as independent as possible, the required decision points and their time requirements were based on the data from the expert analysis and from the tests with the control group in phase 2. This group of younger participants used the interface of the AusweisApp better and without critical errors compared to the elderly users. The experience of these users can therefore be considered to be optimal. In addition, the prediction of the CogTool adds multiple thinking-points automatically where the Keystroke-Level Model considers it.

The theory was that the computation prediction model applied to the optimal user should result lower time requirements for the tasks than the real user tests. This would support the evidence that the users recruited were not optimal and validate the results as being representative. The results for the prototypes for the test column in table 20 were taken from the formative evaluation shown in chapter 4.3.

Table 20 – Comparison of Test and Model time requirements – Selected Prototypes

Interface	Scenario	Test	Model
Original "AusweisApp" N=7	Online Banking	158.3 s	141.2 s
	Video-rental	124.8 s	110.4 s
Prototype 4 N=5	Online Banking	155.2 s	138.3 s
	Video-rental	115.7 s	103.7 s
Prototype 5 N=5	Online Banking	135.3 s	118,0 s
	Video-rental	105.1 s	93.2 s
Prototype 6 N=8	Online Banking	123.6 s	107.4 s
	Video-rental	98.3 s	79.8 s
Prototype 7 N=5	Online Banking	110.8 s	95.9 s
	Video-rental	78.6 s	61.2 s

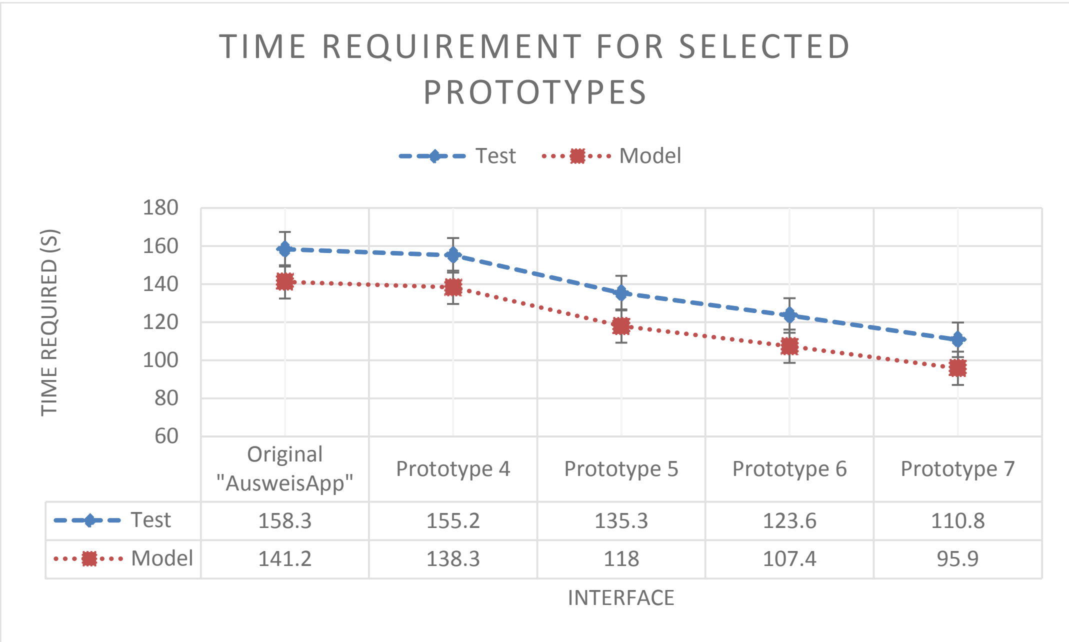


Fig. 30 – Time requirements for selected prototypes – Online-Banking

The results from the tests and the prediction model for the prototypes table 20 shows in the validity of the results. This also illustrates that the sample has less than optimal experience and represents the typical elderly users. It also ensures that the problem described earlier in this chapter did not have a strong impact on the results of the user tests and that they can be used to validate the guideline.

6.5 Compliance and Assessment of the Results Against the Requirements

The final two steps of the IGUAN guideline are the analysis of the compliance to applicable standards (EE3) and the assessment of the design against the requirements (EE4). These functions ensure that the output system is compliant with norms and helps to formulate the final report of the system improvement process.

In case of prototype 7, the ISO 9241-110 is the governing usability standard which the application has to be compliant with. This compliance can be measured by analysing the applicability of the main ergonomic principles of the ISO 9241-110 against the prototype:

- The new system is suitable for the task. This has been verified, as the test participants have completed the set out tasks without critical errors.
- The new system is learnable, the consistency of the system has been improved by the prototyping. This learning effect is also directly observable within the prototypes as the ASQ_{pe} scores for second task were higher than the scores for the first task.
- The conformity with user expectations is met by creating an application, which follows the design language of current systems. The new system is consistent with the design language of the operating system and is therefore easier to learn and use. This is measurable through the ASQ_{pe} and ASQ_{pu} scores. These were significantly higher than the equivalent scores for the initial application.
- The self-descriptiveness is met through the integration of the feedback function in the design domain of the IGUAN.
- Prototype 7 complies with controllability as the user is able to maintain direction and speed over the whole course of the interaction. This has been implemented in the PT domain with the help of the progress indicator.

- The error tolerance of the ISO norm is satisfied through the new arrangement of the visual feedback. Attention of the user is drawn to the better presentation of vital information in certification step.
- The suitability for individualisation is not an essential requirement for the client.

These factors ensure that prototype 7 is compliant with the ISO 9241-110 norm.

The assessment of the new design against the requirements (EE4) has to be accomplished by analysing application against the requirements set out in the first domain.

- *Functional requirements* – The functional components of the software were not redesigned, these are therefore identical with the AusweisApp. The main functionality of prototype 7 is the electronic identification of the user.
- *Design requirements* – The new application fulfils the design requirements set out earlier. The ASQ and the RSME scores indicate that it is pleasant to use. It is compliant with the design language of the operating system it is running on. It looks sophisticated, business-like and modern. The design is usable by the elderly.
- *Application requirements* – The results show that elderly users perceive the system as trustworthy and credible. Font size, feedback and logic of the interface is designed according to the needs of elderly users.

6.6 Discussion and Conclusion

The results from the proof of concept have shown that usability engineering along a scenario based guideline such as IGUAN results in improved acceptance of e-government systems for elderly users. This also implicates that the IGUAN is applicable as a structured approach to the usability improvement process. The proof-of-concept shown in this chapter makes a reproduction of the procedure possible independently from the application and enables a better understanding of the usability improvement process.

The visible performance of a user centred approach such as the IGUAN illustrates the possibilities of e-government, which have not been utilised until now. Other research groups throughout Europe (van Velsen, van der Geest, ter Hedde, & Derks, 2009), (Schedler & Summermatter, 2007) have also come to this conclusion. The general statement that e-government can present major benefits to edlerly, particularly reducing or eliminating difficulties related to the natural ageing process has also been declared by other studies. Compared to Lines et al. (Lines, Ikechi, & Hone, 2007) our research made a further step from evaluation of this statement into development of a possible solution and it offers a tool to translate the evaluated requirements into system design.

The solution shown in our research by introducing the IGUAN guideline is a first step into this direction. The proof-of-concept phase has shown however, that more research in this field is necessary if we want to utilise the full possibilities of e-government for elderly. Our research also built upon the results by Czaja and Lee (Czaja & Lee, 2007) and followed their recommendation to develop guidelines and examples to aid the design process of applications for elderly. They note however that this cohort is not homogeneous and such guidelines might not be enough. The involvement of the elderly into the design process, which was incorporated in the first phase of our research, might be the solution to this problem according to Czaja and Lee.

The connection of these inputs resulted the IGUAN guideline, which connects the recommendations of Czaja and Lee with the conclusions of other research groups about the behaviour of elderly users with e-services. As we haveshown earlier in our research the interface of elderly users and e-government is not fully understood, but we hope that guidelines like the IGUAN can help service providers to offer systems which are accepted by this demographic group.

7. Outlook and the Future of E-government for the Elderly

Our vision is that e-government is not only seen by providers as a possibility to reduce costs, but also as a prospect to offer higher quality services for all citizens. The access to these applications should not be reduced through usability and acceptance barriers; any member of the society should be able to use these systems without regard of age or previous experience. Our research accomplished in Germany and Hungary describes a guideline for the successful implementation of such systems. This guideline is however only a tool, which has to be used by providers in order to achieve a higher acceptance of e-government which truly adds additional value for both the government and the citizens.

It will be challenging but not impossible to offer systems which are perceived by the *"late majority*" and "*laggards*" (Rogers, 1962) as a superior way of communication with the government. The vision for such systems should include an effort for a European Union wide standardisation, which would enable a barrier free access of applications, independent from language, experience or complexity. Upcoming challenges include the utilisation of new technologies for an improvement of services quality. Strategies for optimal usability will need to be adjusted faster and faster as new technology offers not only new ways of interaction, but also paradigm shifts in GUI and human-computer interaction. Touch screens on tablet computers need new ideas of usability compared to "normal" computers with keyboard and mouse input. In the three years of this research, the pace and impact of this paradigm shift could already be observed. Early 2010, when our research was initiated, tablet computers were uncommon and expensive. By 2012, tablets have become a mainstream medium as the technology barrier has been overcome and the systems are becoming more and more sophisticated and affordable. Implementation of e-government systems with touch screen input will require more in depth research. This input concept might bring improvements for elderly users. Schneider and Vetter (Schneider & Vetter, 2008) has shown that elderly users clearly prefer this input method over mouse and keyboard. We therefore expect that the implementation of e-government systems with touchscreen input will increase the acceptability of these services by elderly users. The special issues of this will however need further research, as it is not fully understood how core requirements of e-government systems can be implemented on such devices.

An additional problem, which will be visible in the future of e-government, will be the connection of trust and acceptance. It is a critical commodity, especially for elderly users, when forming communities and implementing new ways of government (Nordfors, Ericson, Lindell, & Lapidus, 2009). Questions about the encouragement of trust and acceptance will have to be answered in the future if e-government shall be successful in maintaining the promise of efficiency.

In the near future, wide acceptance will be needed to reach goals of supranational and multilateral programs like the Europe 2020, and to enhance the e-government penetration in each member state of the European Union to levels seen in Sweden or other Nordic countries. As shown in our research, these systems become increasingly complex and expensive, therefore it is increasingly vital to assess and incorporate user feedback. The assessment of the demand will therefore become one of the most important requirements for governments. Standards or guidelines like the IGUAN developed in our research will be invaluable for this task and will lead to better and more efficient systems built in accordance with the demands of the citizens. Systems with enhanced will transform the G2C communication in the next decade giving citizens a higher service quality and better user experience.

References

Abramson, M., & Morin, T. (2003). IBM Endowment for The Business of Government: E-Government 2003. Lanham: Rowman & Littlefield Publishers, Inc.

Adkisson, H. (2002). Identifying De-Facto Standards for E-Commerce Web Sites. Seattle: University of Washington.

Akatsu, H., Miki, H., & Hosono, N. (2007). Design Principles Based on Cognitive Aging. In J. A. Jacko, Human-Computer Interaction (pp. 3-10). Heidelberg: Springer.

Anderson, J. R., & Lebiere, C. (1998). The atomic components of thought. Mahwah, NJ: Erlbaum.

Arch, A. (2008, May 14). Web Accessibility for Older Users: A Literature Review. Retrieved from W3C: http://www.w3.org/TR/wai-age-literature/

ARD/ZDF-Online-Studie. (2007). Internetnutzung zwischen Pragmatismus und YouTube-Euphorie. Germany: ARD-ZDF-Projektgruppe Multimedia.

Arzi, Y., & Shtub, A. (1997). Learning and forgetting in mental and mechanical tasks: a comparative study. IIE Transactions, 29, 759-768.

Atkins, W. (2004). The Smart Card Report. Elsevier Advanced Technologies: Oxford.

Bailey, B. (2005, August). www.usability.gov. Retrieved Dezember 12, 2011, from Usability.gov: http://www.usability.gov/articles/newsletter/pubs/082005news.html

Barnum, C. M. (2011). Usability Testing Essentials. Burlington, MA: Morgan Kaufmann.

Barnum, C., Bevan, N., Cockton, G., Nielsen, J., Pool, J., & Wixon, D. (2003). The "Magic Number 5": Is It Enough for Web Testing? Conference on Human Factors in Computing Systems 2003 (pp. 698-699). Ft. Lauderdale, FL: ACM.

Batinic, B., & Appel, M. (2008). Medienpsychologie. Heidelberg: Springer.

Baumann, K. (1999). Matrix Evaluation Mtehod for Planned Usability Improvements Based on Customer Feedback. In H.-J. Bullinger, & J. Ziegler, Ergonomics and User Interfaces (pp. 701-706). Mahwah, NJ: Lawrence Erlbaum Associates Inc. Publishers.

Becker, S. A. (2005, February). E-government usability for older adults. Communication of the ACM, 48(2), pp. 102-104.

Bernard, M., Liao, C., & Mills, M. (2001). Determining the Best Online Font for Older Adults'. Retrieved from Software Usability Research Laboratory: http://www.surl.org/usabilitynews/31/fontsr.asp

Bertoa, M. F., & Vallencillo, A. (2010). Usability metrics for software compontents. Retrieved from University of Malaga: http://www.lcc.uma.es/~av/Publicaciones/04/QAOOSE2004.pdf

Beyer, H., & Holtzblatt, K. (1997). Contextual Design: Defining Customer-Centered Systems. San Fransciso: Morgan Kaufmann Publishers.

Birkland, J. L. (2003). Exploring Impacts on Older Adults’ E-Services Usage. Syracuse, NY: Syracuse University.

Birren, J. E., & Schaie, W. K. (1996). Handbook of the Psychology of Aging. San Diego, Ca: Academic Press.

Bødker, S. (2000). Scenarios in user-centred designĐsetting the stage for reflection and action. Interacting with Computers(13), pp. 61-75.

Boehm, B. (1986). A spiral model of software development and enchancement. ACM SIGSOFT Software Engineering Notes, 14-24.

Bortz, J., & Döring, N. (2006). Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Heidelberg: Springer.

Bortz, J., & Lienert, G. A. (2008). Kurzgefasste Statistik für die klinische Forschung. Heidelberg: Springer.

Briand, L., Labiche, Y., & Sauve, A. (2006). Guiding the Application of Design Patterns Based on UML Models. ICSM 2006, Proceedings of the 22nd IEEE International Conference on Software Maintance (pp. 234-243). Washington DC: IEEE Computer Society.

Bruder, C. (2008). Gestaltungsprinzipien für das Training älterer Benutzer elektronischer Geräte. Berlin: Technische Universität Berlin.

Bruder, C., Blessing, L., & Wandke, H. (2007). Training the elderly in the use of electronic devices. In J. Jacko, Human-Computer Interaction (pp. 637-646). Berlin Heiderlberg: Springer.

Capgemini, Rand Europe, IDC, Sogeti and DTi. (2009). Smarter, Faster, Better eGovernement. European Commission.

Card, S. K., Moran, T. P., & Newell, A. (1980). The Keystroke-Level Model for User Performance Time with Interactive Systems. ACM Communications, 23(7), pp. 396-410.

Carter, L., & Belanger, F. (2005). The utilization of egovernment services: Citizen trust, innovation and acceptance factors. Information Systems Journal, 15, pp. 5-25.

Centeno, C., van Bavel, R., Burgelman, J.-C., & Jean-Claude, B. (2005). A Prospective View of e-Government in the European Union. Sevilla: European Commison, Joint Research Centre.

Chadwick-Dias, A., McNulty, M., & Tullis, T. (2003). Web usability and age: how design changes can improve performance. Proceedings of the 2003 conference on Unversal Usability (pp. 30-37). Vancouver: ACM press.

Coolican, H. (2009). Research Methods and Statistics in Psychology. London: Hodder Education.

Coyne, K., & Nielsen, J. (2002). Web Usability for Senior Citizens: 46 Design Guidelines Based Usability Studies with People Age 65 and Older. Nielsen Norman Group Report.

Czaja, S. J., & Lee, C. C. (2007). The impact of aging on access of the technology. Univ Access Inf Soc, 5, pp. 341-349.

Czaja, S. J., & Sharit, J. (1998). Age differences in attitudes toward computers. Journal of Gerontology, 53B, pp. 329-340.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319-339.

de Jong, M., & Lentz, L. (2006). Scenario evaluation of municipal Web sites: Development and use of an expert-focused evaluation tool. Government Information Quarterly, 191-206.

De Jong, M., & Van der Geest, T. (2000). Characterizing Web heruistics. Technical Communication, 311-326.

Deutsche Akademie der Naturforscher Leopoldina. (2009). Gewonnene Jahre Empfelungen der Akademiegruppe Altern in Deutschland. Halle (Saale): Deutsche Akademie der Naturforscher Leopoldina.

Deutsche Akkreditierungsstelle. (2010). Leitfaden Usability. Berlin: Deutsche Akkreditierungsstelle.

Deutsche Rentenversicherung. (2009). Angaben für gesetzl. Altersrenten. Deutsche Rentenversicherung.

Dickinson, A., Eisma, R., & Gregor, P. (2003). Challenging interfaces/redesigning users. CUU '03 Proceedings of the 2003 conference on Universal usability (pp. 61-68). New York, NY: ACM.

Dutke, S., & Reimer, T. (2000). Evaluation of two types of online help for application software. Journal of Computer Assisted Learning, 16, pp. 307-315.

Eller, B. (2009). Usability Engineering in der Anwendungsentwicklung. Wiesbaden: GWV Fachverlage GmbH.

European Parliament and European Council. (2006). Directive 2006/123/EC of the European Parliament and of the Council. Brussels: European Council.

Eurostat. (2011, 12 5). Eurostat. Retrieved from http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&init=1&plugin=1&language=en&pcode=tsdgo330

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Test for correlation and regression analyses. Behavior Research Methods, 41(4), 1149-1160.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3:A flexible statistical power analysis program for the social, behavoral, and biomedical sciences. Behavior Reseach Methods, 39(2), 175-191.

Faulkner, L. (2003). Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. Behavior Reseach Methods, Instruments, & Computers, 35(3), 379-383.

Fisk, A. D., Rogers, W. A., Charness, N., Czaja, S. J., & Sharit, J. (2009). Designing for Older Adults. Boca Raton, FL: CRC Press.

Fittkau & Maaß Consulting. (2009, February 10). www.w3b.org. Retrieved May 23, 2012, from Fittkau & Maaß Consulting: http://www.w3b.org/e-commerce/online-einkaufe-werden-abgebrochen-wenn-gewunschte-zahlungsart-fehlt.html

Fleischmann, U. M. (2008). Inteligenz, Lernen, Denken. In G. G. Oswald W.D., Gerontopsychologie: Grundlagen und klinische Aspekte zur Psychologie des Alterns (pp. 27-43). Wien: Springer Verlag.

Forlizzi, J. (2010). All look same? A comparission of experience design and service design. Interactions of the ACM, 60-62.

Fromm, J. (2010, November). Der neue Personalausweis ist da. DFN Mitteilungen(79), pp. 28-31.

Futó, I. (2007). Electronic Tax Declaration in Hungary. In J. Malkolm, & G. Orthofer, E-Taxation: State & Perspectives (pp. 183-195). Linz: Trauner Verlag.

Gould, J. D., & Lewis, C. (1985, March). Designing for Usability: Key Principles and What Designers Think. Communication of the ACM(3), pp. 300-311.

Granizio, C. J., Yánez, P. L., Ramírez, D. P., & Machado, P. C. (2011). Usability in E-government Sites. 2011 Eighth International Conference on Information Technology: New Generations (pp. 453-458). Las Vegas, NV: IEEE.

Grifoni, P. (2009). Multimodal human computer interaction and pervasive services. New York NY: IGI Global.

Grönlund, Å. (2002). Introduction to Electronic Government: Design, Application and Management. Umea University, Sweden: Umea University, Sweden.

Hackos, J. T., & Redish, J. (1998). User and task analysis for interface design. University of Michigan: Wiley.

Hart, T. A. (2004). Evaluation of Websites for Older Adults: How "Senior-Friedly" Are They?

Hart, T. A., Chaparro, B. S., & Halcomb, C. G. (2008, May-June). Evaluating websites for older adults: adherence to ‘senior-friendly’. Behaviour & Information Technology, 27(3), pp. 191-199.

Hartson, A. W. (2003). Criteira For Evaluating Usability Evaluation Methods. International Journal of Human-Computer Interaction, 145-181.

Hassenzahl, M. (2010). Experience Design. Morgan & Claypool.

Hassenzahl, M., & Sandweg, N. (2004). From Mental Effort to Perceived Usability: Transforming Experiences into Summary Assessments. CHI 2004 (pp. 1283-1286). Wien: ACM.

Hassenzahl, M., & Seewald, F. (2004). VOM KRITISCHEN EREIGNIS ZUM NUTZUNGSPROBLEM: Die Qualitative Analyse in Diagnostischen Usability Tests. In M. Hassenzahl, & M. Peissner, Usability Professionals 2004. Stuttgart: Usability Professionals' Association.

Hinds, P. J. (1999). The Curse of Expertise: The Effects of Expertise and Debiasing Methods on Predictions of Novice Performance. Journal of Experimental Psychology: Applied, 5(2), pp. 205-221.

Hipp, C. (2009). Elderly Friendly HMI: Graphical User Interface forAmbient Assisted Living. Stuttgart: Fraunhofer IAO.

Holzinger, A. (2005, January). Usability Engineering Methods for Software Developers. Communications of the ACM, 1, pp. 71-74.

Holzinger, A., Searle, G., Kleinberger, T., Seffah, A., & Javahery, H. (2008). Investigating Usability Metrics for the Design and Development of Applications for the Elderly. In K. Miesenberger, ICCHP 2008 (pp. 98-105). Heidelberg: Springer.

Hunyadi, L., & Vita, L. (2003). Statisztika Közgazdászoknak. Budapest: Központi Statisztikai Hivatal.

Inforum. (2008). e-Befogadás Magyarországon. Inforum.

International Organization for Standardization. (2010). ISO 9241-210:2010. Geneva: International Organization for Standardization.

Isbister, K., & Schaffer, N. (2008). Game Usability. Burlington MA: Morgan Kaufmann Publishers.

ISO. (1998). Ergonomic requirements for office work with visual display terminals (VDTs) Part 11: Guidance on Usability. ISO.

ISO. (2001). ISO/IEC 9126-1:2001 Software engineering - Product quality - Part 1: Quality model. ISO.

IT Governance Institute. (2007). COBIT 4.1. Rolling Meadows, IL, USA: IT Governance Institute.

Ithaka Consulting. (2009). Digitális Írástudás Akcióterv (DIAT). Budapest: State Chancery - Centre for Electronic Government.

Jacko, & Sears. (2003). The human-computer interaction handbook: fundamentals, evolving technologies. Mahwah, NJ: Lawrence Erlbaum Associates Inc. Publishers.

John, B. E. (2009). CogTool. Retrieved Oktober 5, 2011, from http://cogtool.hcii.cs.cmu.edu/

John, B. M. (2009). CogTool User Guide. Retrieved Oktober 8, 2011, from http://cogtool.hcii.cs.cmu.edu/sites/default/files/CogToolUserGuide_1_1.pdf

Jones, R. E., Mossey, M. E., & Endsley, M. R. (2010). Website Usability Design Principles for the Elderly. In V. Rice, Advances in Understanding Human Performance Neuroergonomics, Human Factors Design, and Special Populations (pp. 833-841). Boca Raton, FL: CRC Press.

Jung, R. (2002). Formale Bewertung der Benutzerkomplexität bildschirmgeschützter Informations- und Unterhaltungssysteme im Kraftfahrzeug. ZMMS Spektrum, 15(VDI).

Kaiser, R. (2009). C++ mit Microsoft Visual C++ 2008. Heidelberg: Springer.

Kantner, L., & Rosenbaum, S. (1997). Usability studies of WWW sites: Heurisitic evaluation vs. laboratory testing. Proceedings SIGDOC Conference (pp. 153-160). Snowbird, Utah.

Kohlhofer, P. (2009). Eine Usability und Accessibility Untersuchung von Styleguide 2.0 E-Government-Formularen in Österreich. Wien.

Konradt, U., Wandke, H., Balazs, B., & Christophersen, T. (2003, May-June). Usability in online shops: scale construction, validation and the influence on the buyers' intention and decision. Behaviour & Information Technology, 22(3), pp. 165-174.

Kotamraju, N. P., & van der Geest, T. M. (2012, March). The tension between user-centred design and e-government. Behaviour & Information Technology, 31(3), 261-273.

Központi Statisztikai Hivatal. (2012, 06 14). A 15–74 éves népesség száma legmagasabb iskolai végzettség szerint, nemenként (1998–). Retrieved from www.ksh.hu: http://www.ksh.hu/docs/hun/xstadat/xstadat_eves/i_qlf015.html

Krug, S. (2010). Rocket Surgery Made Easy. Berkley CA: New Riders.

Kuhlmey, A. (2006). Gesundes Altern - geht das? Berlin: Institut für Medizinische Soziologie - Universitätsmedizin Berlin.

Lathrop, D., & Ruma, L. (2010). Open Government. Sebastopol, California, USA: O'Reilly Media, Inc..

Lenk, K. (1994). Information systems in public administration: from research to design. Informatization in the Public Sector. 3, pp. 307-324.

Lenk, K. (1998). Reform Opportunities Missed: Will the innovative potential of information systems in public administration remain dormant forever? Information, Communication & Society 1, pp. 163-181.

Lenk, K., & Traunmüller, R. (2002). Electronic Government. Aix-en-Provence: Springer.

Lewis, C. H. (1982). Using the "Thinking Aloud" Method in Cognitive Interface Design. IBM.

Lewis, J. R. (1991, January). Psychometric evaluation of an after-scenario questionnaire for computer usability studies: the ASQ. SIGCHI Bulletin, 1, pp. 78-81.

Lewis, J. R. (2004). Sample Size Adequacy for Usability Studies: Current Issues. Boca Raton, FL: IBM Pervasive Computing Division.

Lines, L., Ikechi, O., & Hone, K. S. (2007). Accessing e-Government Services: Design Requirements for the Older User. Universal Accesss in HCI. 4556, pp. 932-940. Berlin: Springer.

Lohr, K. A. (2006). A critical Exploration of Web Usability Guidelines and Design Principles with respect to the special Needs of Elders. Paderborn: Universität Paderborn.

Lumsden, J. (2007). Online-Questionnaire Desing Guidelines. In R. A. Reynolds, R. Woods, & J. D. Baker, Handbook of Research on Electronic Surveys and Measurements (pp. 44-64). Idea Group Inc (IGI).

Maeda, J. (2007). Simplicity - Die zehn Gesetze der Einfachheit. München: Spektrum.

Magyaroszag.hu. (2006, 11 15). Hírközpont - Kormányzati Portál. Retrieved from Magyarorszag.hu: https://hirkozpont.magyarorszag.hu/hirek/ebevallas20061115.html/RatingWindow;jsessionid=8BB6A3D617646F23A73DBB15FD6531D7.portal41?struts.portlet.mode=view&struts.portlet.action=%2FratingPortlet%2FrenderDirect&action=e&windowstate=normal&struts.portlet.eve

Mahn, K. (2006). Online-Kompetenz von Senioren. Hamburg: Diplomica Verlag.

Martin, M., & Kliegel, M. (2005). Psychologische Grundlagen der Gerontologie. Stuttgart: W. Kohlhammer.

McInerney, P., & Li, J. (2002, Jul 1). Progress indication. Retrieved from IBM.com: http://www.ibm.com/developerworks/library/us-progind/

Mead, S. E., Spaulding, R. A., Sit, B. M., & Walker, N. (1997). Effects of Age and Training on World Wide Web Navigation Strategies. Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting, (pp. 152-156).

Mehra, B. (2004). Virtual Communities on the Internet: Social Interactions Across the Digital Divide. Univerisity of Kansas: Univerisity of Kansas.

Meier, A. (2009). eDemocracy & eGovernment - Entwicklungsstufen einer demokratischen Wissensgeselschaft. Heidelberg: Springer.

Mentzinis, P. (2009). Elektronische Bürgerdienste stark nachgefragt. Berlin: Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V.

Merriam, S. B., Caffarella, R. S., & Baumgartner, L. M. (2007). Learning in Adulthood. San Francisco, CA: John Wiley & Sons.

Microsoft. (2010). Windows User Experience Interaction Guidelines. Microsoft.

Molich, R. E. (2004). Comparative usability evaluation (CUE-2). Behavior & Information Technology, 65-74.

Molnar, T., & Nguyen, T. V. (2008). Usability Untersuchung Bürgerportal Brandenburg.de. Potsdam: Universität Potsdam.

Moosbrugger, H., & Kelava, A. (2008). Testtheorie und Fragebogenkonstruktion. Heidelberg: Springer.

Naumann, I., & Hogben, G. (2009). Privacy Features of European eID Card Specifications. Bruxelles: ENISA.

Nebe, K., Zimmermann, D., & Paelke, V. (2008). Integrating Software Engineering and Usability Engineering. In S. Pinder, Advances in Human Computer Interaction (pp. 331-350). Vienna: InTech.

Niehaves, B. (2011). Iceberg ahead: On electronic government research and societal aging. Government Information Quarterly, 28, pp. 310-319.

Nielsen, & Loranger. (2006). Prioritizing Web usability. the University of Michigan: New Riders.

Nielsen, J. (1993). Usability engineering. San Diego: Academic Press.

Nokelainen, P. (2004). www.uta.fi. Retrieved December 5, 2011, from Tampereen yliopisto - University of Tampere: http://www.uta.fi/laitokset/aktk/papers/tech_ped_usability/edmedia2004_pn.pdf

Nordfors, L., Ericson, B., Lindell, H., & Lapidus, J. (2009). eGovernment of Tomorrow Future Scenarios for 2020. Swedish Governmental Agency for Innovation System.

Noveck, B. (2009). Wiki Government. Harrisonburg, Virginia, USA: R.R.Donnelley.

OECD. (2011). E-Government at a Glance. OECD Publishing.

Országos Nyugdíjbiztosítísási Főigazgatóság. (2009). Budapest: Országos Nyugdíjbiztosítísási Főigazgatóság.

Österreichische Computer Geselschaft. (2012). Österreichische Computer Gesellschaft. Retrieved 05 19, 2012, from http://www.ocg.at/ak/edemocracy/wiki2/doku.php?id=projekte:oesterreich:e-government_und_e-voting_in_oesterreich

Oswald, F., & Wahl, H.-W. (2010). Enviromental Perspectives of Ageing. In D. D., & C. Phillipson, The SAGE Handbuch of Social Gerontology. London: ISBN: 978-1-4129-3464-0.

Oswald, W. D., Gatterer, G., & Fleischmann, U. M. (2008). Gerontopsychologie. Heidelberg: Springer.

Park, K. S., & Lim, C. H. (1999). A structured methodology for comparative evaluation of user interface designs using usability criteria and measures. International Journal of Industrial Ergonomics, 23, pp. 379-389.

Petrie, H., & Kheir, O. (2007). The Reletionship between Accessibility and Usability of Websites. CHI 2007 Proceedings - Empirical Studies of Web Interactions, (pp. 397-406). San Jose, CA.

Prensky, M. (2001, October). Digital Natives, Digital Immigrants. (M. U. Press, Ed.) On the Horizon, 9(5).

Ramayah, T., & Ignatius, J. (2002). Impact of Perceived usefulness, Perceived ease of use and Perceived . Penang, Malaysia: Universiti Sains Malaysia.

Richter, M., & Flückinger, M. (2010). Die 7±2 wichtigsten Usability-Methoden. In M. Richter, & M. Flückiger, Usability Engineering kompakt (pp. 21-76). Heidelberg: Spektrum Verlag.

Righi, V., Sayago, S., & Blat, J. (2011). Towards understanding e-Government with older people and designing an inclusive platform with them. International Journal of Public Information Systems, 3, pp. 131-142.

Robinet, A., Picking, R., & Grout, V. (2008). A Framework for Improving User Experience in Ambient Assisted Living. Plymouth: University of Plymouth.

Rogers, E. (1962). Diffusion of innovations. Glencoe: Free Press of Glencoe.

Rosson, M. B., & Carroll, J. M. (2002). Usability engineering: scenario-based development of human-computer interaction. San Francisco CA: Morgan Kaufmann Publishers.

Sayago, S., Sloan, D., & Blat, J. (2011). Everyday use of computer-mediated communication tools and its evaluation over time: An ethnogrpahical study with older people. Interacting with Computers(23), pp. 543-554.

Schedler, K., & Summermatter, L. (2007, April). Customer orientation in electronic government: Motives and effects. Government Information Quarterly, 24(2), pp. 291-311.

Scherer, M. (2008). Senioren im Internet Beispiele für die Überwindung des altersspezifischen Digital Divide. Zürich: Universität Zürich.

Schneider, N., & Vetter, S. (2008). Altersdifferenzierte Adaption der Mensch-Rechner Schnittstelle. IAW SPECTRUM, 3, 1-3.

Seale, J., McCreadie, C., Turner-Smith, A., & Tinker, A. (2002). Older people as partners in assistive technology research: The use of focus groups in the design process. (I. Press, Ed.) Technology and Disability, 14, pp. 21-29.

Sengpiel, M., & Dittberner, D. (2008). The computer literacy scale (CLS) for older adults - development and validation. Mensch & Computer, pp. 7-16.

Sharit, J., Czaja, S. J., Nair, S. N., & Lee, C. C. (2003). The effects of age and environmental support in using telephone voice menu systems. Human Factors, 45, pp. 234-251.

Shneiderman, B. (2003). Promoting universal usability with multi-layer interface design. CUU '03 Proceedings of the 2003 conference on Universal usability (pp. 1-8). New York, NY: ACM.

Snellen, I., & van de Donk, W. (. (1998). Public Administration in an Information Age. Amsterdam: IOS Press.

Sørum, H., Andersen, K. N., & Vatrapu, R. (2012, July). Public websites and human–computer interaction: an empirical study of measurement of website quality and user satisfaction. Behaviour & Information Technology, 31(7), pp. 697-706.

Stapelkamp, T. (2007). Screen und Interfacedesign. Heidelberg: Springer.

Statistisches Bundesamt. (2011). Bildungsstand der Bevölkerung. Wiesbaden: Statistisches Bundesamt.

Statistisches Bundesamt. (2011). Demografischer Wandel in Deutschland. Wiesbaden: Statisches Bundesamt.

Sutcliffe, A. (2010). Designing for User Engagement. Morgan & Claypool.

Tan, W., Liu, D., Bishu, R. R., Muralidhar, A., & Meyer, J. (2001). Design improvements through user testing. Proceedings of the Human Factors and Ergonomics Society, 1181-1185.

Tidwell, J. (2005). Designing Interfaces. Sebastapol, CA: O'Reilly Media Inc.

TNS Infrastest. (2011). (N)ONLINER Atlas 2011. Initiative D21.

Turner, C. W., Nielsen, J., & Lewis, J. R. (2004). Current Issues in the Determination of Usability Test Sample Size. http://drjim.0catch.com/samsize-upa02.pdf.

United Nations. (2010). United nations e-government survey 2010: leveraging e-government at a time of financial and economic crisis. New York: United Nations.

van Velsen, L., van der Geest, T., ter Hedde, M., & Derks, W. (2009, July). Requirements engineering for e-Government services: A citizen-centric approach and case study. Government Information Quarterly, 26(3), pp. 477-489.

Venkatesh, V., & Bala, H. (2008). Technology Acceptance Model 3 and a Research Agenda on Interventions. Decision Sciences, pp. 273-315.

Wandke, H., & Jörn, H. (1999). Zum Navigationsverhalten von Anfängern im World Wide Web. Zeitschrift für Arbeits- und Organisationspsychologie, 43(1), S. 46-54.

Wandke, H., Sengpiel, M., & Sönksen, M. (2012, October). Myths About Older People's Use of Information and Communication Technology. Gerontology, 58(6), pp. 564-570.

Wherton, J., & Prendergast, D. (2009). The Build Bridges Project: Involving Older Adults in the Design of a Communication Technology to Support Peer-to-Peer Social Engagement. In A. Holzinger, & K. Miesenberger, HCI and Usability for e-Inclusion (pp. 111-134). Heidelberg: Springer.

Wind, M., & Kröger, D. (2006). Handbuch IT in der Verwaltung. Heidelberg: Springer Verlag.

Wixon, D., Holzblatt, K., & Knox, S. (1990). Contextual design: an emergent view of system design. Conferece on Human Factors in Computer System archive: Proceedings of the SIGCHI conference on Human Factors in Computing Systems (pp. 329-336). Seattle WA: Empowering People.

World Health Organisation. (2012). WHO - Definition of an older or elderly person. Retrieved from WHO: http://www.who.int/healthinfo/survey/ageingdefnolder/en/index.html

Wright, P., Blythe, M., & McCarthy, J. (2006). User Experience and the Idea of Design in HCI. Interactive Systems, Design, Specification and Verification Lecture Notes in Computer Science. 3941, pp. 1-14. Newcastle upon Tyne: Springer.

Wright, T. P. (1936). Factors affecting the cost of airplanes. Journal of Aeronautical Sciences, 3, pp. 122-128.

Yeh, S., Rubin, E. S., Hounshell, D. A., & Taylor, M. R. (2009). Uncertainties in Technology Experience Curves for Integrated Assessment Models. Retrieved 06 22, 2012, from berkeley.edu: http://gspp.berkeley.edu/academics/faculty/docs/mtaylor/Yeh,%20Rubin-Uncertainties%20in%20Tech%20ExpCurves%20for%20IA%20Models%20-%20IJETP%20accepted%2005.pdf

Zentrum für Sichere Informationstechnologie Austria. (2008). Verbreitung der Bürgerkarte. A-SIT.

Zhao, F. (2013). An empirical study of cultural dimensions and e-government development: implications of the findings and stragies. Behaviour and Information Technology, 32(3), pp. 294-306.

Ziefle, M., & Bay, S. (2003). Mental models of a cellular phone menu: Comparing older and younger novice users. In B. S., & M. Dunlop, Mobile human computer interaction. Berlin, Heidleberg: Springer.

Zijstra, F. R. (1993). Efficency in work behaviour. A design approach for modern tools. Delft: Delft University of Technology.

Appendix A – The questionnaires

A1 – Phase 1 Questionnaire – Germany

Testperson: ………… (Nummer) – **Bitte keine Namen eintragen**

Geschlecht: Weiblich: ☐ Männlich: ☐

Alter:

Ausbildung:

Primäre Ausbildung:
Grundschule oder Vergleichbar

Sekundäre Ausbildung:
Abitur, Berufsausbildung, Fachausbildung oder Vergleichbar:

Tertiäre Ausbildung:
Universitätsabschluss, Fachhochschulababschluss, Berufsakademie oder Vergleichbar:

Erfahrung mit dem Umgang mit Computern:

Keine:
„Ich habe noch nie einen Computer benutzt“

Niedrig:
„Ich benutze einen Computer sehr selten“

Mittel:
„Ich benutze einen Computer manchmal und komme damit bei einfachen Dingen, wie im Internet stöbern, oder eine Email schreiben, zu Recht.“

Hoch:
„Ich benutze einen Computer oft, und kann damit auch komplexe Aufgaben erledigen.“

Wohnort:

Ländlich: Vorort: Innenstadt:

Computer zu Hause: JA: NEIN:

Ich habe schon eine E-government Applikation genutzt (z.B.: ELSTER):

JA: NEIN:

Das Szenario:

Level 0 – Offline

Ich empfand diese Vorgehensweise als Einfach:

JA: NEIN:

Welche Wartezeit ist für Sie in einem Amt noch angemessen?:

Keine:
Weniger als 20 Minuten:
Weniger als 30 Minuten:
Weniger als 40 Minuten:
Weniger als 50 Minuten:
Weniger als 1 Stunde:
Mehr als eine Stunde: (Bitte konkrete Angabe)

Haben Sie bei diesem Verfahren genügend Informationen über die Leistungen erhalten?

JA: NEIN:

Level 1 - Informationen im Internet verfügbar

Diese Vorgehensweise ist einfacher als die vorherige:

JA: NEIN:

Die Neuerungen bei dieser Vorgehensweise sind nützlich:

JA: NEIN:

Durch die neuen Komponenten wurde der Vorgang schneller.

JA: NEIN:

Ich würde so ein System nutzen:

JA: NEIN:

Level 2 – Online Formulare

Diese Vorgehensweise ist einfacher als die vorherige:

JA: NEIN:

Die Neuerungen bei dieser Vorgehensweise sind nützlich:

JA: NEIN:

Durch die neuen Komponenten wurde der Vorgang schneller.

JA: NEIN:

Ich würde so ein System nutzen:

JA: NEIN:

Level 3-4 – Online Kommunikation

Diese Vorgehensweise ist einfacher als die vorherige:

JA: NEIN:

Die Neuerungen bei dieser Vorgehensweise sind nützlich:

JA: NEIN:

Durch die neuen Komponenten wurde der Vorgang schneller.

JA: NEIN:

Ich würde so ein System nutzen:

JA: NEIN:

Der persönliche Kontakt im Amt ist für mich wichtig:

JA, ich brauche den persönlichen Kontakt NEIN, der Online Vorgang ist besser

A2 – Phase 1 Questionnaire – Hungary

Tesztszemély: .... (Szám) – **Kérem ne írja be a nevét**

Nem: Nő: ☐ Férfi: ☐

Kor:

Iskolai Végzettség:

Alapfokú:
Általános Iskola vagy hasonló:

Középfokú:
Érettségi, Szakma vagy hasonló:

Felsőfokú:
Egyetem, Főiskola vagy hasonló:

Számítógép használat, tapasztalat:

Nincs:
„Soha nem használtam számítógépet"

Alacsony:
„Nagyon ritkán használok számítógépet"

Közepes:
„Néha használok számítógépet, egyszerűbb feladatokat, internet böngészés, email írás segítség nélkül képes vagyok elvégezni"

Magas:
„Gyakran használok számítógépet, komplexebb feladatok sem jelentenek problémát"

Lakhely: Falu, Község: Előváros: Belváros:

Van számítógépem otthon (jelenleg):

van: nincs:

Használtam már elektronikus kormányzati szolgáltatást, (pl. elektronikus adóbevallás:

igen: nem:

A Szenárió:

0. Szint - Offline

Egyszerű volt az ügyintézés:

igen: nem:

Mennyi az Önnek elfogadható maximális várakozási idő?:

Semmi:
Kevesebb mint 20 perc:
Kevesebb mint 30 perc:
Kevesebb mint 40 perc:
Kevesebb mint 50 perc:
Kevesebb mint 1 óra:
Több mint 1 óra:

Elegendő volt a szolgáltatással kapcsolatos információ (A szenárión belül)

Elég információt kaptam a hivataltól: Kevés információt kaptam a hivataltól:

1. Szint – online elérthető információ

Ez a megközelítés egyszerűbb ügyintézést biztosít, mint az előző:

igen: nem:

Az ebben a megközelítésben megjelent újdonságok hasznosak:

igen: nem:

Az új komponensek felgyorsítják az ügyintézést.

igen: nem:

Használnék ilyen rendszert:

igen: nem:

2. Szint – online formanyomtatványok

Ez a megközelítés egyszerűbb ügyintézést biztosít, mint az előző:

igen: nem:

Az ebben a megközelítésben megjelent újdonságok hasznosak:

igen: nem:

Az új komponensek felgyorsítják az ügyintézést.

igen: nem:

Használnék ilyen rendszert:

igen: nem:

3-4. Szint – Teljesen online ügyintézés

Ez a megközelítés egyszerűbb ügyintézést biztosít, mint az előző:

igen: nem:

Az ebben a megközelítésben megjelent újdonságok hasznosak:

igen: nem:

Az új komponensek felgyorsítják az ügyintézést.

igen: nem:

Használnék ilyen rendszert:

igen: nem:

A személyes kapcsolatot az ügyintézésnél elengedhetetlennek tartom:

igen, a személyes kapcsolat fontos: nem, az online ügyintézés előnyösebb:

A3 – Phase 2 Questionnaire – Germany

Erfassungs- und Fragebogen – Nutzbarkeitstest des Bürgerclients

Promotionsforschungsprojekt: Nutzbarkeit von electronic Government Systemen für die alternde Bevölkerung

Tamas Molnar

10.09.2010

Humboldt-Universität zu Berlin

Teil A – Erfassungsbogen – Wird vom Testleiter ausgefüllt

Datum: ....
Testperson Nr.: ....

Geschlecht:
Alter:

Aufgabe 1:

Aufgabe erfolgreich durchgeführt:

JA:
Mit externer Hilfe:
NEIN:

Anzahl der externen Hilfen:

einfach – würde auch ohne Hilfe weiterkommen

mittel komplex – würde ohne Hilfe nach trial and error weiterkommen

sehr komplex – würde ohne Hilfe den Test abbrechen

Benötigte Zeit zur Durchführung der Aufgabe:

Beanspruchung für den Nutzer

Testperson wurde von der Aufgabe nicht sichtbar beansprucht:

Testperson wurde von der Aufgabe auf ein Minimum beansprucht, die Aufgabe erschien schwer:

Testperson wurde von der Aufgabe stark beansprucht, die Aufgabe erschien als sehr schwer:

Testperson wurde von der Aufgabe stark frustriert, hat aufgegeben:

Aufgabe 2:

Anzahl der externen Hilfen:

einfach – würde auch ohne Hilfe weiterkommen

mittel komplex – würde ohne Hilfe nach trial and error weiterkommen

sehr komplex – würde ohne Hilfe den Test abbrechen

Benötigte Zeit zur Durchführung der Aufgabe:

Beanspruchung für den Nutzer

Testperson wurde von der Aufgabe nicht sichtbar beansprucht:

Testperson wurde von der Aufgabe auf ein Minimum beansprucht, die Aufgabe erschien schwer:

Testperson wurde von der Aufgabe stark beansprucht, die Aufgabe erschien als sehr schwer:

Testperson wurde von der Aufgabe stark frustriert, hat aufgegeben:

Erlernbarkeit der Aufgabe

Die Testperson war schneller und sichtbar weniger beansprucht bei der 2. Aufgabe:

JA NEIN

Teil B – Fragebogen – Wird von der Testperson ausgefüllt

Haben Sie Zugang zu einen Computer mit Internetanschluss (Desktop, Laptop, iPad oder ähnliche Geräte inbegriffen) in Ihrem Wohnumfeld?

JA NEIN

Seit wie vielen Jahren benutzen Sie einen Computer?

Wie viele Stunden pro Woche nutzen Sie einen Computer?
Wie oft Nutzen Sie einen Computer für folgende Tätigkeiten? (Bitte ankreuzen)

	nie	selten	gelegentlich	oft
Textverarbeitung				
Tabellen				
Präsentationen				
Bildbearbeitung				
Programmieren				
Computerspiele				
Online Spiele				
Email				
Internet Suche				
Online-Einkäufe				
Online-Banking				
Social Networks				
E-government				

Aufgabe 1:

Bitte benoten Sie die von Ihnen empfundene Schwierigkeit der Aufgabe mit 1 bis 5. (1. Sehr Einfach...5.Sehr Schwer)

Bitte benoten Sie die von Ihnen empfundenen Nutzen der Applikation: (1. Sehr Nützlich...5. Überflüssig)

Ich bin mit der Zeit in der ich die Aufgabe gelöst habe zufrieden. (1. Stimme zu..5. Stimme nicht zu)

Ich bin mit den erhaltenen Informationen zufrieden. (1. Stimme zu..5. Stimme nicht zu)

Würden Sie eine solche Anwendung nutzen?

JA:
Möglicherweise, wenn diese besser gestaltet sein würde:
NEIN:

Was hat für Sie den größeren Nutzen bei solchen Anwendungen: (bitte benoten)

Persönlicher Kontakt: Schnelle Durchführbarkeit:

Aufgabe 2:

Empfanden Sie diese Aufgabe einfacher als die 1.?

JA: NEIN:

Bitte benoten Sie die von Ihnen empfundene Schwierigkeit der Aufgabe mit 1 bis 5. (1. Sehr Einfach...5.Sehr Schwer)

Bitte benoten Sie die von Ihnen empfundenen Nutzen der Applikation: (1. Sehr Nützlich...5. Überflüssig)

Ich bin mit der Zeit in der ich die Aufgabe gelöst habe zufrieden.(1. Stimme zu..5. Stimme nicht zu)

Ich bin mit den erhaltenen Informationen zufrieden. (1. Stimme zu..5. Stimme nicht zu)

Würden Sie eine solche Anwendung nutzen?

JA:
Möglicherweise, wenn diese besser gestaltet sein würde:
NEIN:

Was hat für Sie den größeren Nutzen bei solchen Anwendungen: (bitte benoten)

Persönlicher Kontakt: Schnelle Durchführbarkeit:

A4 – Phase 2 Questionnaire – Hungary

Kérdőív – Használhatósági teszt elektronikus személyigazolvány kliens

PhD. Kutatás: Elektronikus közigazgatási szolgáltatások használhatóságának vizsgálata az öregedő társadalomban

Molnár Tamás
29.10.2010

Humboldt-Universität zu Berlin

A rész- Erfassungsbogen– Wird vom Testleiter ausgefüllt[9]

Datum: ....
Testperson Nr.: ....

Geschlecht:
Alter:

Aufgabe 1:

Aufgabe erfolgreich durchgeführt:

JA:
Mit externer Hilfe:
NEIN:

Anzahl der externen Hilfen:

einfach – würde auch ohne Hilfe weiterkommen

mittel komplex – würde ohne Hilfe nach trial and error weiterkommen

sehr komplex – würde ohne Hilfe den Test abbrechen

Benötigte Zeit zur Durchführung der Aufgabe:

Beanspruchung für den Nutzer

[9] No need for translation into Hungarian, as it was not given out to the test candidates

Testperson wurde von der Aufgabe nicht sichtbar beansprucht:

Testperson wurde von er Aufgabe auf ein Minimum beansprucht, die Aufgabe erschien schwer:

Testperson wurde von der Aufgabe stark beansprucht, die Aufgabe erschien als sehr schwer:

Testperson wurde von der Aufgabe stark frustriert, hat aufgegeben:

Aufgabe 2:

Aufgabe erfolgreich durchgeführt:

JA:
Mit externer Hilfe:
NEIN:

Anzahl der externen Hilfen:

einfach – würde auch ohne Hilfe weiterkommen

mittel komplex – würde ohne Hilfe nach trial and error weiterkommen

sehr komplex – würde ohne Hilfe den Test abbrechen

Benötigte Zeit zur Durchführung der Aufgabe:

Beanspruchung für den Nutzer

Testperson wurde von der Aufgabe nicht sichtbar beansprucht:

Testperson wurde von er Aufgabe auf ein Minimum beansprucht, die Aufgabe erschien schwer:

Testperson wurde von der Aufgabe stark beansprucht, die Aufgabe erschien als sehr schwer:

Testperson wurde von der Aufgabe stark frustriert, hat aufgegeben:

Erlernbarkeit der Aufgabe

Testperson war schneller und sichtbar weniger beanspruht bei der 2. Aufgabe:

JA NEIN

B rész– Kérdőív – A teszt résztvevője tölti ki

Van számítógép internet-hozzáféréssel az ön lakókörnyezetében? (Desktop, Laptop, iPad vagy hasonló eszközök)?

IGEN NEM

Hány éve használ ön számítógépet?

Hány órát használ ön egy héten átlagosan számítógépet?
Milyen gyakran használ ön számítógépet a következő tevékenységekre? (Kérem X-el jelezze)

	soha	ritkán	néha	gyakran
Szövegszerkesztés				
Táblázatkezelés				
Előadások elkészítése				
Képszerkesztés				
Programozás				
Számítógépes játékok				
Online játékok				
Email				
Internetes keresés				
Internetes vásárlás				
Internet banking				
Social network				
Elektronikus kormányzat				

1. Feladat:

Kérem osztályozza a feladatot nehézsége szerint:
(5. Nagyon egyszerű...1.Nagyon bonyolult)

Kérem osztályozza a feladat hasznosságát:
(5. Nagyon hasznos...1. Felesleges)
A megoldás időigényével elégedett vagyok:
(5. Teljesen egyetértek..1. Nem értek egyet)

A rendelkezésre álló információkkal meg vagyok elégedve:
(5. Teljesen egyetértek..1. Nem értek egyet)

Használnék egy ilyen alkalmazást.

IGEN:
Talán, ha jobban lenne megvalósítva:
NEM:

Mi fontosabb ön számára egy ilyen elektronikus alkalmazásnál: (kérem ossztáyozza)
Személyes ügyintézés: Lehető leggyorsabb ügyintézés:

2. Feladat:

Egyszerűbbnek találtam ezt a feladatot, mint az 1.?

IGEN NEM

Kérem osztályozza a feladatot nehézsége szerint:
(5. Nagyon egyszerű...1.Nagyon bonyolult)

Kérem osztályozza a feladat hasznosságát:
(5. Nagyon hasznos...1. Felesleges)

A megoldás időigényével elégedett vagyok:
(5. Teljesen egyetértek..1. Nem értek egyet)

A rendelkezésre álló információkkal meg vagyok elégedve:
(5. Teljesen egyetértek..1. Nem értek egyet)

Használnék egy ilyen alkalmazást.
IGEN:
Talán, ha jobban lenne megvalósítva:
NEM:

Mi fontosabb ön számára egy ilyen elektronikus alkalmazásnál: (kérem osztályozza)
Személyes ügyintézés: Lehető leggyorsabb ügyintézés:

Appendix B – Results

B1 – Understand the Context – Interviews

Table B1 – Control variable gender – Phase 1

Gender	Germany	Hungary	Sum
Female	24	13	**37**
Male	21	12	**33**
Sum	45	25	**Sum = 70**

Table B2 – Control variable education

Education	Germany	Hungary	Sum
Primary or None	4	2	**6**
Secondary	26	11	**37**
Tertiary	15	12	**27**
Sum	45	25	**70**

Table B3 – Control variable previous experience with e-government systems

Previous Experience	Germany	Hungary	Sum
Low or None	26	18	**44**
Medium	17	7	**24**
High	2	0	**2**
Sum	45	25	**70**

Table B4 – Control variables; Mean (Standard Deviation) – Phase 1

Mean	Germany	Hungary	Sum
Age (in Years)	70.4 (4.4)	69.5 (4.3)	**70.1 (4.4)**
Previous Experience[10]	1.47 (0.59)	1.28 (0.46)	**1.4 (0.55)**
Computer in Living Environment	0.58 (0.5)	0.36 (0.49)	**0.5 (0.5)**
Has used e-government	0.29 (0.46)	0.24 (0.44)	**0.27 (0.45)**

[10] 3 stage scale (1 = low, 3 high experience)

Correlation and t-test

$$r = \frac{cov(x,y)}{s_x * s_y}$$

$$t = \frac{r * \sqrt{n-2}}{\sqrt{1-r^2}}$$

Table B5 – ASQ Factor Scores – Mean (Standard Deviation)

	ASQ Factor	Germany	Hungary	Sum
Level 1	Utility	44 (0.15)	23 (0.28)	**67 (0.2)**
	Time Gain	42 (0.25)	25 (0)	**67 (0.2)**
	Ease of Use	43 (0.21)	23 (0.28)	**66 (0.23)**
Level 2	Utility	26 (0.5)	19 (0.44)	**45 (0.48)**
	Time Gain	33 (0.45)	18 (0.46)	**51 (0.45)**
	Ease of Use	30 (0.48)	20 (0.41)	**50 (0.46)**
Level 3/4	Utility	13 (0.46)	5 (0.41)	**18 (0.44)**
	Time Gain	25 (0.5)	11 (0.51)	**36 (0.5)**
	Ease of Use	15 (0.48)	9 (0.49)	**24 (0.48)**

B2 – Specify user requirements – Heuristics and Tests

Table B6 – Control variable gender – Phase 2

Gender	Germany	Hungary	Control Group	Sum
Female	22	13	10	**45**
Male	23	17	10	**50**
Sum	45	30	20	**Sum = 95**

Table B7 - Control variables; Mean (Standard Deviation) – Phase 2

Mean	Germany	Hungary	Control Group	Sum
Age (in Years)	70.62 (6.29)	68.27 (6.35)	27.95 (3.95)	**69.68 (6.38)**[11]
Previous Experience (in Years)	4.39 (3.55)	3.25 (1.48)	N/A[12]	**4,05 (3.1)**
Computer in Living Environment	0.62 (0.49)	0.4 (0.5)	N/A	**0.53 (0.5)**
Computer Use/Week	3.38 (3.67)	1.87 (2.47)	N/A	**2.77 (3.31)**

Table B8 – Results from the CLS Part A – Phase 2

CLS Part A[13]	Never	Rarely	Occasionally	Frequently
Internet Search	12	1	29	33
Email	15	8	28	24
Text processing	24	34	15	2
Online-Banking	36	23	16	0
Online-Shopping	53	9	8	5
E-government	62	7	3	3
Photo Editing	58	5	12	0
Computer Gaming	57	6	7	5

Table B9 – Successful competition of the scenarios

Success		Germany	Hungary	Control Group	Sum
Scenario 1	Abort	0	0	0	**0**
	Success	1	0	6	**7**
	Success with Support	44	30	14	**88**
Scenario 2	Abort	0	0	0	**0**
	Success	30	22	19	**71**
	Success with Support	15	8	1	**24**

[11] Without Control Group
[12] Not measured, would have falsified the interpretation of the data
[13] Without Control Group

Table B10 – Absolute number of error types (Mean, Standard Deviation) – Phase 2 – Scenario 1

Errors	Germany	Hungary	Control Group
Minor (1)	3.21 (2.15)	3.8 (2.14)	1.75 (1.28)
Medium (2)	1.57 (1.28)	1.63 (1.3)	1.11 (0.78)
Critical (3)	1.82 (0.75)	1.8 (0.84)	0.2 (0)

Table B11 – ASQ Scores (Mean, Standard Deviation) - Phase 2 – Both Scenarios

ASQ		Germany	Hungary	Control Group	Sum[14]
Scenario 1	Utility	3.36 (1.26)	3.27 (1.31)	3.14 (1.21)	**2.3 (1.26)**
	Time Gain	2,02 (0.62)	1.77 (0.57)	1.4 (0.5)	**1.81 (0.62)**
	Ease of Use	1.71 (1.16)	1.97 (0.89)	1.1 (0.31)	**1.66 (1)**
Scenario 2	Utility	3.29 (0.97)	3.87 (0.9)	3.75 (0.85)	**1.62 (0.6)**
	Time Gain	1.71 (0.59)	1.8 (0.61)	1.15 (0.37)	**2.36 (1)**
	Ease of Use	2 (0.71)	2.1 (0.71)	1.73 (0.79)	**1.98 (0.73)**

B3 – Produce design solutions – Proof of Concept of the IGUAN

Table B12 – Control variable gender – Phase 3 – Summative Evaluation Prototype 7

Gender	Germany	Hungary	Sum
Female	22	15	**37**
Male	18	20	**38**
Sum	40	35	**Sum = 75**

Table B13 - Control variables; Mean (Standard Deviation) – Phase 3 – Summative Evaluation Prototype 7

Mean	Germany	Hungary	Sum
Age (in Years)	69.7 (3.97)	69.82 (5.62)	**69.76 (4.78)**
Previous Experience (in Years)	2.53 (3.5)	2.72 (3.94)	**2.8 (3.6)**
Computer in Living Environment	0.47 (0.51)	0.36 (0.49)	**0.46 (0.5)**
Computer Use/Week	3.22 (3.71)	2.44 (3.28)	**2.94 (3.56)**

[14] Including Control Group Scores

Table B14 – Results from the CLS Part A – Phase 3 – Summative Evaluation – Prototype 7

CLS Part A	Never	Rarely	Occasionally	Frequently
Internet Search	14	3	40	18
Email	16	10	31	18
Text processing	19	20	20	16
Online-Banking	42	30	3	0
Online-Shopping	52	14	5	4
E-government	69	3	3	0
Photo Editing	64	3	7	1
Computer Gaming	71	3	1	0

Table B15 – Control variable Age – Prototyping (Mean, Standard Deviation) – Both scenarios

Prototype	No. of Participants	Age (M)	SD
Prototype 1	5	72.60 years	5.74
Prototype 2	13	70.60 years	5.77
Prototype 3.1	14	71.25 years	5,03
Prototype 3.2	13	70.22 years	5.40
Prototype 4.1	5	72.60 years	3.72
Prototype 4.2	5	71.80 years	3.86
Prototype 5	8	71.00 years	3.60
Prototype 6.1	8	71.00 years	3.60
Prototype 6.5	7	70.20 years	2.92
Prototype 7	7	70.80 years	3.76

Table B16 - ASQ_{pe} Scores – Prototyping (Mean, Standard Deviation) – Scenario 1

Prototype	ASQ_{pe}Score	SD
AusweisApp 4.1	1.84	1.92
Prototype 1	1.59	1.19
Prototype 2	1.92	0.49
Prototype 3.1/3.2	1.97	0.49
Prototype 4.1	2.40	0.54
Prototype 4.2	2.20	0.83
Prototype 5.1	1.40	0.71
Prototype 5.2	2.20	0.86
Prototype 5.3	2.60	1.01
Prototype 6.1	2.67	0.89
Prototype 6.5	2.80	0.83
Prototype 7	3.00	0.70

Table B17 - ASQ_{pe} Scores – Separated/Integrated user guidance (Mean, Standard Deviation) – Scenario 1

Prototype	Participants	ASQ Score for ease-of-use	
		M	SD
Separated	5	2.20	0.837
Integrated	9	1.75	0.782

Table B18 - ASQ_{pe} Scores – PIN Input field (Mean, Standard Deviation) – Scenario 1

N=8	Mean ASQ ease-of-use	Mean Time Requirement
Sequential	1.4	194 s
"Calculator"	2.2	143 s
"Telephone"	2.6	130 s

Table B19 - Absolute number of error types (Mean, Standard Deviation) – Phase 3 – Scenario 1

Errors	Germany	Hungary
Minor (1)	1.42 (0.99)	1.35 (0.77)
Medium (2)	0.5 (0.79)	0.7 (0.71)
Critical (3)	0 (0)	0 (0)

Table B20 – RSME scores for the applications – Both scenarios

RSME score	AusweisApp 4.1	Prototype 7
Scenario 1	58.45	41.98
Scenario 2	39.72	31.10

Appendix C – Proof of concept – Hypotheses and results (SPSS)

C1 – Correlation between Problems and Nationality

Table C1 – Correlation between nationality and number of errors

Korrelationen

		Minor	Medium	Critical	Nationality
Minor	Korrelation nach Pearson	1	,123	,029	-,120
	Signifikanz (2-seitig)		,292	,804	,303
	N	75	75	75	75
Medium	Korrelation nach Pearson	,123	1	-,099	-,017
	Signifikanz (2-seitig)	,292		,398	,888
	N	75	75	75	75
Critical	Korrelation nach Pearson	,029	-,099	1	,005
	Signifikanz (2-seitig)	,804	,398		,965
	N	75	75	75	75
Nationality	Korrelation nach Pearson	-,120	-,017	,005	1
	Signifikanz (2-seitig)	,303	,888	,965	
	N	75	75	75	75

C2 - T-test – Prototype 2 – The certificate problem – ASQ scores

Table C2A/B – T-Test Prototype 2 - ASQ Scores

Gruppenstatistiken

	SW_Version	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
ASQ_Pe	A41D	12	1.50	.522	.151
	P2D	13	1.92	.494	.137

Test bei unabhängigen Stichproben

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
ASQ_Pe	Varianzen sind gleich	3.572	,071	-2,083	23	,049	-.423	.203	-.843	-,003
	Varianzen sind nicht gleich			-2,078	22.559	,049	-.423	.204	-.845	-,001

C3 - T-test – Prototype 2 – The certificate problem - Errors

Table C3A/B – T-Test Prototype 2 - Errors

Gruppenstatistiken

	SW_Version	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Error_3	A41D	5	1.60	1.140	.510
	P2D	5	.20	.447	.200

Test bei unabhängigen Stichproben

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
Error_3	Varianzen sind gleich	3.881	,084	2.556	8	,034	1.400	.548	.137	2.663
	Varianzen sind nicht gleich			2.556	5.202	,049	1.400	.548	,008	2.792

C4 - T-test – Prototype 3 – The progress indicator

Table C4A/B – T-Test Prototype 3

Gruppenstatistiken

	SW_Version	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
ASQ_Pe	P31D	14	1.79	.426	.114
	P32D	13	2.15	.555	.154

Test bei unabhängigen Stichproben

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
ASQ_Pe	Varianzen sind gleich	.197	.661	-1.943	25	,063	-.368	.189	-.758	,022
	Varianzen sind nicht gleich			-1.924	22.505	,067	-.368	.191	-.764	,028

C5 - T-test – Prototype 4.0 – Separation of logical steps

Table C5A/B – T-Test Prototype 4

Gruppenstatistiken

	SW_Version	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
ASQ_Pe	P4D	5	2.20	.837	.374
	P32D	9	2.11	.782	.261

Test bei unabhängigen Stichproben

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
ASQ_Pe	Varianzen sind gleich	,035	.855	.199	12	.846	,089	.446	-.884	1.062
	Varianzen sind nicht gleich			.195	7.893	.850	,089	.456	-.965	1.143

C6 - T-test – Prototype 4.1 – Implementing selection boxes

Table C6A/B – T-Test Prototype 4.1

Gruppenstatistiken

	SW_Version	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
ASQ_Pe	P4D	5	2.20	.837	.374
	P41D	5	2.40	.548	.245

Test bei unabhängigen Stichproben

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
ASQ_Pe	Varianzen sind gleich	.640	.447	-.447	8	.667	-.200	.447	-1.231	.831
	Varianzen sind nicht gleich			-.447	6.897	.668	-.200	.447	-1.261	.861

C7 - F-test – Prototype 7 – GUI Final

$$\boldsymbol{F = \frac{s_2^2}{s_1^2} = \frac{(\frac{1}{n_2} - 1)\sum_{i=1}^{n_2}(x_{2i} - \bar{x}_2)^2}{(\frac{1}{n_1} - 1)\sum_{i=1}^{n_1}(x_{1i} - \bar{x}_1)^2}}$$

$$F = 5.74, F_{crit}(6.6) = 4.28$$

C8 - T-test – Summary of the improvements

Table C7A/B – T-Test Prototype 7

Gruppenstatistiken

	SW_Version	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
ASQ_Pe	A41D	75	1.60	.805	,093
	P7D	7	3,00	.577	.218

Test bei unabhängigen Stichproben

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
ASQ_Pe	Varianzen sind gleich	3.436	,067	-4.481	80	,000	-1.400	.312	-2,022	-.778
	Varianzen sind nicht gleich			-5.902	8.355	,000	-1.400	.237	-1.943	-.857

C9 – T-test – summative evaluation prototype 7 against AusweisApp

Gruppenstatistiken

	SW_Version	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
ASQ_Pe	A41	75	1,60	,805	,093
	P7DS	75	3,00	,885	,102

Test bei unabhängigen Stichproben

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
ASQ_Pe	Varianzen sind gleich	1,019	,314	-10,130	148	,000	-1,400	,138	-1,673	-1,127
	Varianzen sind nicht gleich			-10,130	146,694	,000	-1,400	,138	-1,673	-1,127